

**STATE ASSESSMENT SYSTEMS: THE USE OF RESULTS, IMPLICATIONS AND
TRENDS**

SANDRA ZÁKIA SOUSA

sanzakia@usp.br

ROMUALDO PORTELA DE OLIVEIRA

romualdo@usp.br

Professors of the College of Education of the University of São Paulo

Versão para o inglês: Thaddeus Gregory Blanchete

ABSTRACT

This article presents the results of research that analyzes the assessment systems implemented by five Brazilian states (Bahia, Ceará, Minas Gerais, Paraná and São Paulo), considering their characteristics in 2005-2007. Taking into account their specificity, an attempt is made to explore how the results have been used for formulating and implementing educational policies, as well as their potential for becoming a milestone in educational policy that effectively intervenes in the management of school systems and schools. It was observed that the systems tend to have similar characteristics, their main point of reference being the Basic Education Assessment System, on national level. As for using the results, their impact is at best tenuous, or even non-existent, even though, in the words of their proponents, it is preceded by powerful rhetoric, as expressed in the literature about its potential. However, from an examination of the systems carried out by the managers themselves, it is possible to identify a concern with making them more effective. So, even in this study initiatives have already been identified that seek to transform the use of the data obtained into a management tool, as the ones that propose offering bonuses to teachers and staff.

*EDUCATIONAL POLICIES – STUDENTS EVALUATION – EDUCATIONAL
ADMINISTRATION – SAEB*

From the 1990s on, school system evaluation became central to public education policy in Brazil, being recommended and promoted by several international agencies, the Federal Ministry of Education and several State Education Secretariats as an especially effective tool for the improvement of basic and higher education systems. Following these recommendations, several different levels modalities of education began to be evaluated by the government under the presumption that this would produce better quality education (Sousa, 2001, p.90).

Assessment's possible contributions to the improvement of education have generally been discussed following two types of arguments:

- a. The first considers the internal logic of the evaluation processes, examining their principles and methods, seeking to make clear the chosen criteria and focuses and taking these as a reference in discussing the possibilities and limitations of the practices under analysis. The assessment processes' potential for improving education would be present in the intrinsic logic of its own design (Oliveira, 2008);
- b. The second argument also looks at how assessment results are used to delineate educational policy, examining the validity, relevance and the opportunity the adopted initiatives have as a consequence of the knowledge obtained through the assessment process.

In Brazil, debate regarding federal and state government initiative has tended to focus on program principles and guidelines, seeking to explain their potential or limitations and the inherent risks of their implementation and sometimes even exploring the concepts of educational quality that are implicit in various government programs. Investigations which propose to explore the use of assessment results in the various decision-making processes throughout the system of education are still rare (Sousa, 1997) and research along these lines is thus timely and useful. The institutionalization of assessment systems at both the state and federal level has public policy implications for education which involve institutions, curricula, teaching professionals and students, indicating a redefinition of the State's role in the educational field.

From 1995 on, the implementation of assessment processes has intensified in Brazil and the view that the State should increase its role in improving education in our country – an intrinsic part of large scale assessment – has assumed an increasing importance in debates over educational policy. During this period, international agencies began to directly stimulate and influence the outlines of the evaluation system proposals, showing a clear interest in the efficacy of external investments in education. The Basic Education Evaluation System – SAEB – implemented in 1990 on the national level, began to receive partial financing from the World Bank and consequently, suffered transformations, such as changes in its agents, conceptual influences, research and international evaluation techniques (Bonamino, 2002). SAEB's consolidation stimulated other assessment proposals at the state government level, and these initiatives began to be central in the undertaken policies.

This article presents the results of research financed by the State of São Paulo Research Foundation (Fundação de Amparo à Pesquisa do Estado de São Paulo – Fapesp) undertaken between 2005 and 2007 with the goal of characterizing and comparing the assessment systems implemented by 5 states (Bahia, Ceará, Minas Gerais, Paraná e São Paulo) in an attempt to understand their presuppositions and characteristics, as well as describe how their results have informed educational policy formulation and implementation. Our research is based on the understanding that the assessment systems must be analyzed in a way that seeks to go beyond the superficial level of their proposals and analytical instruments. Our goal here is to verify the potential these systems have to affect educational policy and to effectively intervene in schools. The need for research of this type has been pointed out by Barretto and Pinto (2001) and also by Sousa (2002), given that all these authors have shown that few studies exist regarding the impacts and results of assessment systems. Barretto and Pinto, mentioning the few works which exist regarding assessment systems, claim that:

...[there exists] a preponderance of official discourses in these works, principally when it comes to explaining the characteristics and goals of permanent assessment systems which might contribute to improving school quality [...] The most systematic studies undertaken in Brazil to date by the assessment systems themselves – Saeb, Paemg, Saesp – seem to be closed in on themselves. There's no attempt to link these studies to other research on school and educational quality and they do not show any evidence of accumulated reflection on the topics they attempt to analyze. (Barretto, Pinto, 2001, p. 61)

One must also take into account the symptomatic fact that these assessment systems emerge at the moment in which there was a growing questioning of the role the State played in capitalist societies, directed against the State's capacity to manage the services with which it had been historically associated¹ (Sousa et al., 2000; Sousa, Oliveira, 2003). This was more than an academic debate and it resulted in an appreciation of the nature of the State's role in unleashing significant changes in public policy, particularly through the privatization of established services and activities in the public sphere. The discussion of assessment systems took in a wide variety of technical questions and had political implications, given that it was intimately linked to public educational policy.

These assessment systems were not isolated elements in the national and international public educational spheres, and have lately scene an impressive degree of development. In the field of basic education, SAEB² stands out, followed by the National Mid-Level Educational Exam (Exame Nacional do Ensino Médio – Enem) and the National Course Examination (Exame Nacional de Cursos – ENC³). This tendency towards assessment has also gained space on the international stage, either within national systems of education or through procedures that seek to establish comparisons between countries, such as the International Student Evaluation Program (Programa de Avaliação Internacional de Estudantes – PISA⁴).

¹ This affirmation may need to be used in a relative way over the next few years, depending upon what definitions become hegemonic regarding the recent subprime crisis, whose global consequences are not yet clear.

² Instituted in 1990 by the Ministry of Sports and Education, this system undertakes assessments based on a sample of schools in all states, carried out every two years. From 2005 on, it consisted of two initiatives: the National Basic Education Evaluation (Avaliação Nacional de Educação Básica – ANEB) and the National School Productivity Evaluation (Avaliação Nacional do Rendimento Escolar – Anresc). The first is undertaken via sampling in each state school. It focalizes on the systems of education's management, maintaining the basic characteristics of the SAEB. The second system focuses on each individual school.

³ Though they differed in terms of procedures and methods, both ENEM and ENC were to provide indicators regarding mid-level and higher education, respectively, covering a wide range of schools throughout the nation. They exclude primary education (which is still largely dependent upon a very restrictive concept of social work) and post-graduate courses (which are subject to another sort of evaluation carried out by the Coordinating Council for the Improvement of Higher Level Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES).

⁴ PISA evaluates 15 year old students independent of their placement in grades and is applied by the Organization Economic Cooperation and Development (OECD) among member states. The system, however,

Patrícia Broadfoot emphasizes that the dominant matrix is strongly related to a vision in which assessment systems disseminate a managerial system that is strongly influenced by corporate managerial logic:

Perhaps even more important than this increasingly effective control, however is the growing association of educational administration in both countries – England and France – with a corporate management approach. Such an approach is likely to disguise the essentially political nature of educational goals – in the ideology of scientific rationality. In this event, value judgment appears as merely administrative decisions dictated by rationality and the goal of maximizing efficiency. It seems probable that effective educational control implies the existence of a social order ready to concur in educational goals. The way in which assessment procedures help to bring this about will perhaps prove ultimately more significant than their role in imposing such goals, for the requirements of mass testing are such as to require a considerable measure of agreement over educational objects. It may well be that the testing technology overwhelms the initial, sensitive identification of educational goals, reinforcing the long deplored tendency for assessment to reduce curricula goals to what can be measured and bringing a stage nearer the predominance technocratic ideal of managerialism. (Broadfoot, 1996, p. 123-124).

In this perspective, the results utilization of large scale assessment processes faces an intense ideological debate. On the one hand, expectations exist that such processes allow for system management anchored in an instrumentalist rationality which will generate greater efficiency. On the other, there is the ideology that school-based educational processes should be autonomous and self-evaluating.

Given this debate, the issue of using the assessment's results acquires a new dimension: to what degree has the implementation of these systems created a greater self-examination in the system as a whole and in individual schools in particular? Better yet, how have these systems been able to provoke changes in the logic of formulating and implementing educational policies and scholastic practices?

The literature regarding evaluation systems has focused on the regulatory action of the State and its effects along the lines laid out by Madaus:

has a policy of widening its purview to students of other nations, which explains the involvement of Brazilian students in the exams undertaken in 2000. (OECD, 2001, 2001b).

1. The power of tests and examinations to affect individuals, institutions, curriculum. or instruction is a perceptual phenomenon: if students, teachers. or administrators believe that the results of an examination are important, it matters very little whether this is really true or false-the effect is produced by what individuals perceive to be the case.
2. The more any quantitative social indicator is used for social decision making, the more likely it will be to distort and corrupt the social processes it is intended to monitor.
3. If important decisions are presumed to be related to test results, then teachers will teach to the test.
4. In every setting where a high-stakes test operates, a tradition of past exams develops, which eventually de facto defines the curriculum.
5. Teachers pay particular attention to the form of the questions on a high-stakes test (for example, short answer, essay, multiple-choice) and adjust their instruction accordingly.
6. When test results are the sole or even partial arbiter of future educational or life choices, society tends to treat test results as the major goal of schooling rather than as a useful but fallible indicator of achievement.
7. A high-stakes test transfers control over the curriculum to the agency which sets or controls the exam. (Madaus, 1988, p. 94)

Though its limitations are recognized, the quantification of educational quality has been presented as the driving motivation behind many assessment system initiatives which depend above all on the students' performance, either measured via a census or via a sample from the population under study. Along with this perspective generally comes the claim that these assessment practices themselves induce higher quality education. These points, however, are quite controversial and it is difficult to operationally define educational quality, in part because of its social and historical conditioning factors:

...it's necessary to reiterate the fact that assessing educational quality is not limited to simply verifying scholastic production, which is simply one moment in which quality becomes manifest. Students' performance in research regarding the quality of education is better understood and interpreted when one takes into consideration information regarding the kind of teaching the students receive, the procedures they encounter in the classroom and the school and also the environmental characteristics of students' families which determine their behavior. (Vianna, 1990, p. 99)

These concerns gain in relevance when we realize that the main large scale assessments - SAEB and the Brazil Exam – use student's school performance as the main

measure for their analyses of educational quality (Sousa, 2001, p. 88). To say this is not to judge these systems: it is to simply point out the limitations of these procedures which might create a patina of “scientificism” that does not adequately support the policies derived from said assessments (Perrenoud, 1998). Furthermore, such practices do not favor the consolidation of a “true assessment system”, something which demands a wide variety of initiatives that:

...are not easily understood by society and, maybe for this reason, has less appeal to decision and policy-makers. Census-based single-factor assessment systems cause greater emotional and political impact, and one can compare the Big Exam and SAEB in this sense. However, the problem is that not everything that provokes emotions and political impact can make good on its promises. (Franco, 2004, p.61-62).

In the literature under investigation here, there is a large number of works which deal with the assessment processes currently used in the United States and in which these concerns are often repeated (Heubert, Hauser, 1999; Kohn, 2000; Orfield, Kornhaber, 2001; Levin, 2001; Mcneil, 2000; Wilde, 2002, among others). One also finds similar concerns in those works which deal with European assessment systems, such as the following comments by Roggero:

By understanding the educational system or organization to be a complex system which is capable of producing a self-ecoreorganization, we become conscious that this process is dependent upon the system’s history (re), its relationships with the surrounding environment (eco) and its internal identity (self). All assessment activities should thus take into consideration these essential dimensions which we are unable to evaluate via simple performance indicators. (Roggero, 2002, p. 43)

The literature shows that, while assessment initiatives are growing throughout the world, at the same time new technical and political questions are emerging that demand to be included in educational policy and in the research projects that can reveal trends, support critiques and demonstrate paths to the future. One indicator of the generalization of assessment processes in the world is an issue of *Revista de Educación* magazine, published in 200 by Spain’s Ministry of Education and which was dedicated to assessment systems:

One of the concerns expressed by educators of any country is the need to understand how their educational system works in practice, beyond the designs and intentions of legislators. On the other hand, society in general is ever more demanding regarding how educational systems work and what markers of progress they have achieved. Because of this, governments *institutionalize* and make instrumental mechanisms and *assessment processes* in receiving ever more information regarding their educational systems and to pass this information along to the citizenry. (p. 7, our emphasis).

If it's true that the importance of school and school system evaluation has generalized; however, there is still little agreement regarding "what should be assessed" and "how it should be assessed". Obviously, there are linkages between the motives which have created greater institutionalization of evaluation systems and the ways in which assessments are undertaken (Oliveira; Sousa, 1999). The adoption of certain evaluative processes signals what kind of modifications is desired within the educational system. Even though the "winners" and "losers" in the assessment process might not be "awarded" or "punished", the simple fact of assessment itself indicates that an ideal situation has been defined and that all others will be measured in accordance with this.

To undertake a given assessment process is to attempt to verify how distant one is from the idealized situation and thus be able to define elements, which can be modified to move the system in the desired direction. Assessment can as well permit the emergence of proposals to redirect or transform the situation under evaluation. It is, in fact, impossible to think of assessment without also thinking of modification. The most important question thus perhaps becomes how to create involvement that pushes members of an institution to engage themselves in the process of transformation. The criteria and patterns can easily simplify complex problems by using indicators that have the potential to provoke change, but which won't necessarily induce the idealized situation. This observation indicates the need for a more careful reflection regarding the uses (and abuses!) of the Basic Education Development Index (Índice de Desenvolvimento da Educação Básica – Ideb). (Oliveira, 2007).

In the case of state-level assessment systems, we must ask questions regarding their goals, procedures and consequences and how these impact upon the quality of education as decreed by educational policy formed within state governments. Quality, after all...

...is not a "given", it does not exist "in and of itself". It refers to an axiological question: what are the values of those who produce the qualitative assessment? The emergence of assessment criteria does not occur disconnected from the positions, beliefs, worldview and social practices of those who do the

assessing. It is a concept that is born within the philosophical, political and social perspectives of the assessors. The focus and criteria adopted by an assessment process thus reveal the axiological options of the people who participate in the process. (Sousa, 1997, p.26).

In concentrating on the uses of the results of assessment systems, the present study investigate not only what has been done, but also inquires after the consequences of these acts.

METHODOLOGY

In the first phase of our research, we consolidated information regarding the assessment systems of the states of Bahia, Ceará, Minas Gerais, Paraná and São Paulo which was available on the websites of their respective Education Secretariats and in documents produced by these institutions. In some cases, we also looked at information furnished by professionals involved in state government, who responded to our consultations via e-mail. Seeking to comprehend how assessment proposals were inserted into state education policies, we gathered information about these proposals, with an eye to discovering the directives and priorities established by the various governments and the place reserved for assessment in these. We thus consulted documents produced by the several Education Secretariats, as well as academic research and in particular theses and dissertations that took these policies as their subject of investigation.

The period delineated for this research varied from state to state. We took as our starting mark for documentary research the year in which each state implemented its assessment system and tried to collect information up until 2006, where possible. This first phase of the research resulted in a preliminary description of the assessment systems which was later complemented by information collected through fieldwork during the second phase of the investigation. This second phase involved semi-structured interviews whose main objective was to discover how assessment results have been utilized and to understand the importance attributed to assessment by education managers and technicians in formulating and implementing educational policy. The interviews also sought to complement or clarify data regarding the characteristics of the various state educational proposals in terms of their principles, breadth and procedures.

All information was then systematized and analyzed for evidence pertaining to the meanings and uses that assessment has acquired in educational policy. Beginning with the mapping of each state and taking as a reference the set of discourses displayed by education managers and technicians, we have proceeded to a comparative analysis of the states. This, in turn, permits us to indicate trends and perspectives which are based upon the reports of the actual practices of assessment within the educational systems studied.

ASSESSMENT SYSTEMS: PECULIARITIES AND TRAJECTORIES

The state-level assessment systems under analysis here began to be implemented in the early 1990s. The most recent one was organized in Bahia in 1999. Minas Gerais was one of the first states to organize its own assessment system in 1992. The state's current system, however, was reorganized along the lines of that of the state of Ceará in 2000. Ceará's assessment system dates back to the beginning of the 1980s.

The goals of assessment, as declared in the official literature, tend to affirm the expectation that evaluation will help with decision-making in order to improve the educational system at several different levels. Generally, it's hoped that assessment results will be appropriated by managers and the Education Secretariats' central and regional teams, as well as by the schools. Mention is also made in some states' literature of hopes that the educational community as a whole – including students and parents – will make use of the results obtained by the schools. In other words, the literature is replete with comments that assessment should illuminate the path to (re)formulating educational policy via directives, priorities and goals, focalizing attention on management, curricula and infrastructure as well on the proposals and practices of the schools themselves, particularly with regards to curriculum development. In particular, in 1995 Paraná established the objectives of its system as “discovering schools' performance with regards to organization, management and links to the surrounding community”. In 2003, Ceará sought to “implement the formation of school councils, in order to make them able to competently develop their own assessment processes”.

The main indicator that assessment systems have used to judge the quality of education has been students' proficiency at various applied tests. Much mention is made about the need to collect information regarding the students and schools. This information, however, has not been linked to the results achieved by students on their tests. In other words,

there is no indication that the information collected by assessment systems has been used to interpret test scores in a contextualized fashion in an attempt to identify factors which may be associated with students' proficiency levels. It was also observed that literature often announced the intention to create a "culture of assessment" in the educational systems, a concept that is not explained or discussed in such a way as to make its meaning clear. Even the analysis of the procedures utilized in the implementation of the assessment process doesn't give any clue which would allow us to guess what said "culture" could possibly be.

With regards to objectives, in 1992 Ceará established the goal of "creating a quality index in every school" and Paraná decided it wanted "to establish goals for each school". Establishing goals based on assessment results, while not explicitly stated in other states' documents, was an objective cited by several of our interviewees. There was a definite tendency to associate assessment results with specific commitments that would be taken on by the schools and by other institutions within the educational system. This configuration was most clearly explicit in the way in which Minas Gerais conducted its assessment activities.

The application of school performance tests in the states researched has been undertaken on a census-based. There was one exception to this general rule in São Paulo in 2002, however: while the tests were census-based for the schools they were based on a sampling of individual students in each school. Assessment systems have also gone beyond the schools of the state system, being applied to any municipal and private schools which want to take part in the assessment process. Municipal schools joined the assessment process in Minas Gerais and in Ceará. In São Paulo, Bahia and Paraná, private and municipal schools were invited to participate. In all states studied.

There was a strict delimitation with regards to what grades and courses were to be the focus of assessment. Assessment tests have generally been organized on annual basis, though in Paraná and Bahia they have been held every two years. Generally, fourth through eighth grade students and high school students are assessed for Portuguese and Math Paraná also assesses Sciences, Physics, Chemistry and Biology while São Paulo had Reading, Writing and Math exams in the last assessments which we analyzed.

Regarding the construction of the examinations themselves, documents indicate (with differing levels of detail) the references used, the steps taken to elaborate the instruments utilized and the people responsible for the activities proposed in the chronograms established

by the Education Secretariat. Below, we give some examples of elements which permit us to describe tendencies at play in the assessment initiatives of each state under study.

Bahia's Education Secretariat counted on the participation of professors from both private and public schools and based the guidelines for the elaboration of its tests on national curricula parameters as well as the books and didactic practices understood to be generally used in the state by the participating professors. These guidelines were then validated by specialists, who evaluated the set of proposed descriptors with regards to value and pertinence. The choice of which items would actually be included in the tests was left up to the Ministry of Education's assessment team. In Ceará and São Paulo, the test questions were elaborated based upon the respective states' curricula and the national curricula parameters. Specialists in the various areas of knowledge to be tested then revised the questions. Paraná and Minas Gerais structured their guidelines based on the curricula then being used by the state schools, the national curricula parameters and also SAEB's guidelines. In these states, the assessment network's professors and the Education Secretariat's technicians participated in establishing the guidelines, together with outside consultants.

The above information regarding the confection of the exams allows us to suppose that these tests are more or less similar, given that they used common references in their elaboration: the national curricula parameters and SAEB. We must still discover, however, if there are specific differences in the state curricula which set them apart from the national parameters to a large enough degree to create significantly different guidelines in each state. We also need to discover to what degree the participation of professors in these elaborations (under the auspices of linking the guidelines to the proposals then in place in various schools) resulted in proficiency expectations which were different than those expected by SAEB. Questionnaires were also utilized by the assessment systems in order to collect information about the students. Some states also made use of survey instruments to collect information regarding schools and their professors and employees.

Judging by the available information, all the assessment systems took SAEB as their reference and sought to link their proficiency scales to this reference matrix. Some of the programs used three parameters Item Response Theory (IRT) as an analytic tool. This was a specific characteristic of the Ceará program in 1998 and 2002, which used Classic Average Theory as well. This second tool is mentioned specifically by the São Paulo program for the years 2000 and 2001. São Paulo also claims to have used Item Response Theory in 2000, but

gives no further details. In the case of Bahia, we have no information for 1999 and 2000, and the scale originally constructed for 2001 does not directly correspond to that of SAEB. Instead, an equivalency scale was constructed based on “social judgment” with the aid of the American Institute for Research (AIR). According to the material collected, the objective of this scale “was to permit associations such as: average score ‘x’ in the assessment of skills corresponding to an average level localized on the SAEB scale” (Bahia, 2003, p.71). In Minas Gerais in 2000, 2001 and 2002, three parameter ITR was used. Paraná only mentions conducting a “statistical analysis” without adding any further details.

The states presumed that their results, once divulged, would be utilized by different institutional levels and actors within the systems of education. However, during our research, we were not able to obtain access to current and updated assessments data, which indicates that assessment results have not been published according to the planned schedule – in other words, with the same frequency in which the systems conduct their tests.

Paraná presumed that school principals and the municipal education secretariats would divulge students’ results. There was also provision for something called a “pedagogical analysis of the assessment results” which would present considerations and suggestions to professors and the presidents of the local Parent-Teacher Associations based upon the test results. There was also to be a “school report card” which would divulge a simplified version of the results for each school and which would be passed out to students’ parents.

São Paulo planned to publish students’ results on the website of the Education Secretariat and also in regionally distributed reports which would discuss the assessment with schools. Bahia and Minas Gerais also planned to distribute reports throughout the networks’ schools. These would contain general information regarding results on a state level, at the municipal level and also with reference to each particular school and would be followed by a pedagogical analysis of the tests. The same procedure was to be followed in Ceará, which also was to undertake regional seminars in order to discuss the results.

With regards to how the states hoped the assessment results would be used, the most commonly reiterated point was that the assessment would aid in the continued education and professionalization of the teaching staff by improving pedagogical practices. Our interviewees also indicated that assessment results were considered in the planning of training activities for the professors, with emphasis being placed upon the content that students demonstrated difficulties in mastering.

In the case of Ceará, assessment results were associated with prizes. From 2002 on, the New Millennium School: Ceará Quality in Elementary Education prize was given to those schools whose students scored the highest in Ceará Permanent System for Elementary Education Assessment (Sistema Permanente de Avaliação da Educação Básica do Ceará – SPAECE) exam.

In São Paulo in 2001, the results of the State of São Paulo Student Performance Evaluation System (Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo – SARESP) were utilized by the Education Secretariat to determine which students should pass on to the next grade and which should be held back.

In general, our sources indicate a great degree of similarity in the guidelines adopted by the assessment systems. One must point out, however, the initiative undertaken by the Ceará Education Secretariat to integrate SPAECE (a large scale assessment program) with the Institutional Assessment project which is implementing school self-assessment undertaken by all segments of the educational community at all levels. Thus, from 2000 on, SPAECE has integrated both large scale assessment and the institutional assessment process that had earlier been conducted separately from SPAECE. Considering the characteristics of the assessment guidelines adopted by the state, Ceará's assessment system seems to be the most innovative of the five under analysis, given that it incorporates different dimensions, flows, actors and outlooks.

Paraná has struck out in a different direction from Ceará, however, a trajectory which we were able to confirm in our fieldwork. The state has decided to stop large scale assessment and proceed solely with a system of institutional self-evaluation, which will seek to analyze activities at all levels of the educational system.

The guidelines for Bahia's External Assessment Project also have an interesting specific characteristic, at least according to the information released by the state's Education Secretariat (SEC) on its website (<http://www.aval.ufba.br/projeto.asp>). Bahia wants to integrate the activities of its *Educate to Win* strategic program with its Assessment Project. EtW is composed of two large scale assessment systems – Performance Assessment and Learning Assessment – which fulfill different and specific roles. However, according to our informants, this initiative has still not gotten off the ground

Study of the documents pertaining to the state assessment systems allows us to suppose that the implementation of these systems is a tendency which has already

consolidated itself within the field of Brazilian education. Initially, it involved only the state public schools but it has since been extended to the municipal schools and (on a voluntary basis) top private schools. However, direct contact with the managers of the state assessment systems allowed us to see that if these systems have integrated their initiatives, they also felt a need to rethink their scope in order to become more relevant to the management and promotion of quality education. There's also the feeling that certain initiatives are being duplicated at the state level. In fact, since 2005, the Brazilian federal government began to assess the systems of education in a census-based, through the 'Brazil Exam'.

The state assessment systems studied in our research presented many common characteristics in their original design. Interviews with managers showed us, however, that a movement has begun at the level of the state Education Secretariats to seek out assessment perspectives which have greater potential impacts in terms of subsidizing policy-making and inducing the production of idealized results.

Many new initiatives at the state level thus seem to present their own particularities. One of these refers to the implantation of institutional self-assessment systems based on the understanding that such systems generate (among other things) greater involvement with and commitment to improving the quality of education. The experience of Ceará in this sense, which is not recent and which has been maintained within the greater structure of large scale assessment, and that of Paraná, implemented in 2005, are examples of this.

Another type of initiative, one which strikes out in a completely different direction, intensifies the connections between assessment and consequences. This linkage can occur at any point in the process, from the elaboration of a specific action plan for negotiation, accompaniment and control up to giving out prizes for certain types of results. Minas Gerais, for example, has implanted a work dynamic which has intensified the relationship between assessment, planning and control within the public school network. Meanwhile, Ceará has followed this path even more decisively by giving out prizes in accordance with assessment results, although this sort of thing has also begun to appear in the other State Education Secretariats. Many new proposals suggest that schools and teachers should be awarded according to their assessment results.

Managers have many expectations with regards to the contributions that assessment systems may make to education. These range from consequences in formulating and executing educational policies to influencing in school curricula and management. We can

definitely claim that we are currently passing through a fertile period when it comes to the construction of new assessment dynamics.

Apparently, Education Secretariats are seeking assessment alternatives that can contribute to policy formation and management. It must be recognized that a large quantity of information is accumulating do to the already-implemented assessment programs and that this is a new fact for the Brazilian education system as a whole. This information is still primarily focused on student's performance, however. Data regarding the greater social and economic context of the schools themselves is collected, but so far has been underutilized as a potential resource for explaining the differences in school performance which have been identified within the system. There is also no evidence that this information is being used in a consistent manner as part of a process of systemic assessment, although this is somewhat understandable due to the relative newness of the currently employed systems. Finally, its worth emphasizing the fact that SAEB has been elected as something of a central reference for the managers involved in setting up and running state-level assessment systems, both in terms of its design and as a parameter against which other alternatives are proposed.

ASSESSMENT SYSTEMS: TENDENCIES AND PERSPECTIVES

Our research has permitted us to identify characteristics generally present in Brazil's education assessment systems. We discuss some of the more important of these below, noting how they present challenges which must be confronted not only by the systems' managers, but also by those researchers who've chosen the educational assessment field as their focus of study, particularly in terms of its connections to educational policy-making.

TRUSTWORTHINESS OF RESULTS

Over the last few years, the technical capacities needed to carry out large scale assessment have improved greatly in Brazil, particularly with regards to average theory. A good example of this is the generalization of the use of Item Response Theory, both with two and three parameters, in the elaboration and testing of exam questions. Another example has been the consolidation of institutions that have the capacity to carry out such assessments,

even though the World Bank is correct in pointing out that there is, as of yet, no significant market for this sort of service in Brazil.

Certain problems have persisted, however, and these threaten the trustworthiness of assessment results in our country. One of these difficulties has to do with practices related to the applying assessment instruments. It's still common, for example, for a school's professors to apply the tests themselves, even though its results will have an impact upon their salaries or may determine the bonuses or awards which the school receives. A further problem has to do with the political interest that Education Ministers and other representatives of the state government have in achieving positive assessment results, as these are generally seen as affirming the with of current education policy.

CHANGING PROPOSALS

Only one of the five states studied – Ceará – has continued on with its original proposal and plans for an assessment system, characterizing changes which have been made since the system's founding as "improvements". The other four states have not continued with their original systems, either due to changes in the state government or due to the federal government's plans for evaluating all schools via the Brazil Exam.

Changes in plans due to changing governments are a general characteristic of Brazilian public administration and education is no exception to this. When there is a change in the Minister of Education, it is expected that early education initiatives tend to be abandoned, too. In the specific case of assessment systems, this generally has resulted in the suppression of the system being used by the earlier administration or by altering the system's goals, procedures, or uses, corrupting its integrity and making it impossible to evaluate its results as part of a historical series.

It must also be recognized that the construction of the state assessment systems was, in large part, due to federal incentives and the creation of SAEB as a model. Also, many World Bank financing projects insisted upon the development of state-level assessment systems as part of their clauses. In this sense, then, many of the state systems were technically created to complement SAEB and this means that their characteristics (curricula tested, analysis

methodology, test topics, etc.) were to originally be compatible with that system or were made to be compatible as time went on⁵.

When the federal government unilaterally decided to begin census-based testing with the Brazil Exam, two types of problems quickly arose. In the first place, the states began to question the validity of maintaining their own expensive and logistically complex assessment systems in order to carry out the same tasks as the new federal system. Secondly, the Brazil Exam, taken together with state exams, increased test frequency to the point where it became impossible to adequately analyze and absorb the results of one assessment and implement suggested transformations throughout the educational system before another was already taking place.

We can infer that some of the managers of the systems under study here would have already abandoned their own state systems if they had confidence that the Brazil Exam would indeed be applied on a regular basis. The current arrangement also seems to be the result of a structural problem. The only reason to promote census-based testing instead of sample-based testing (which is perfectly adequate for information collection purposes and is significantly cheaper) is in order to manage the entire school system in accordance with the assessment process through the inclusion of each school and student in the results. Given that the states have not yet reached an agreement with the federal government to use the Brazil Exam as an educational system management tool, there's an additional problem with regards to the incorporation of its results into state-level policy, aside from the above-mentioned issue with the results coming too quickly to be adequately utilized⁶.

If these initiatives become better established and a more stable administrative and technical culture coheres with assessment being understood to be a general State policy (and not the policy of a specific government), wider use of exam results will probably be achieved.

⁵ There has been an exchange between the state-level systems under study. This means that some initiatives have been undertaken due to criticisms of unsuccessful experiences in other educational systems. Likewise, successful experiments in other states have been adopted and absorbed. A concrete example of this is the state of Minas Gerais' use of Bahia's experience of certifying school principals.

⁶ One must also take into consideration, however, the fact that the Brazil Exam's results have been integrated into the IDEB which, together with the School Census, constitute a reference for the definition of goals to be achieved by the public school systems by 2021.

EMPHASIS ON MAKING SCHOOLS RESPONSIBLE

In the formulation of assessment system goals, one often hears mention of the expectation that the results of the system will be used to subsidize the decisions made at all levels in the school network, from central managers down to the schools themselves. In some cases, reference is even made to the idea that the school community itself will make use of the assessment results. In seeking concrete examples of how results have been used, we have verified that most initiatives taken have generally situated the schools as the biggest user of test results, being understood to be responsible for these. The system managers responsible for initiatives of this sort, however, recognize that the schools face difficulties in comprehending and even reading the results produced by the assessment system. Because of this, many principals have taken steps to have the results translated into a more accessible language. This can include explanatory seminars and the production of syntheses which are distributed to the schools.

It is not, however, part of the culture and dynamic of schools to base their planning on assessment results. Because of this salient fact, the initiatives referred to above have yet to make much impact on school dynamics. In the eyes of education professionals, it seems that the assessment which is understood to be the most significant is an evaluation of the individual student's learning process conducted by each professor. Attempts such as that made by the state of São Paulo to utilize the results of the assessment system as a defining element of students' scholastic progression have been understood by schools as a loss of their autonomy and power. Thus we find a difference between what managers believe can be done with assessment results (making schools responsible for their use) and what schools actually do with the results (practically unable to understand them, let alone use them).

The use of assessment system results by the managers themselves is practically non-existent. On this level, one observes the tendency to understand such results as simple indicators (which should be juxtaposed with others such as dropout statistics, student repetition statistics, etc.) and not as something that should be used to inform specific policies. The management of the system is based upon bureaucratic structures and not oriented by the results of school performance. The largest investment – in human and financial resources – is made in terms of directly producing information and not in producing systems able to stimulate the use of assessment results. The notion of accountability – both in terms of giving

an ‘account’ to society and in terms of "making responsible" – is as yet little explored. It seems that even the parents of the assessed students interact very superficially with the assessment process and its results.

PRIVITIZATION OF ACTIVITIES

The Education Secretariats studied maintain teams responsible for assessing their systems at different points in their on-paper structure. Among the five states studied, however, only São Paulo and Ceará actually have consolidated teams provided with the technical skills necessary to clearly interpret assessment results and their possible uses. Technicians who’ve long been a part of the teaching system are integrated into these teams and this has guaranteed an institutional comprehension of the states’ assessment systems, even if these have long become transformed by political decisions. Minas Gerais has established a dynamic in which technical knowledge regarding assessment has been incorporated directly into the upper echelons of the Educational Ministry and it is hoped that this will become an important factor in implementing new education policies. In the other two Education Secretariats, personnel in the assessment structures have changed with great frequency. This has degraded capacities to the point where, in one of the states studied, no member of the assessment team was even able to open the database furnished by the company which had been contracted to carry out the elaboration, application and analysis of the assessment exams.

Independent of questions regarding the assessment teams’ configuration, there’s the issue of the logistic complexity of implementing the assessment process itself, which has led to the contracting of institutions and/or companies to conduct the tests. This creates a further series of limitations. On the one hand, the assessment teams are no longer able to control the assessment system’s design and become dependent upon the timeline of the contracted institution. More importantly, the teams have to trust in the quality of the results presented by the hired institution as well as the procedures they utilize. On the other hand, by contracting testing and evaluation out to third parties, the assessment teams never learn themselves how to conduct such activities. The competencies of the state team responsible for the formulation and implementation of educational policy are thus not in harmony with those of the technicians contracted to carry out the assessment procedures themselves. Depending on the type of negotiations established between the Education Secretariat and the institutions actually

doing the testing and analysis, the power to appropriate assessment results and translate them into quality-improvement activities may be quite limited. An extreme case in this sense was identified in a system in which the results of each school were turned over to that school alone, without being distributed to the Education Secretariat's central team.

Discontinuities in the assessment initiatives have been aggravated by the delays in ratifying contracts with the testing companies and institutions. Aside from the regular delays in public contracting processes such as these (which include respecting the present in the contracts and which originate in Brazilian legislation), there are also criteria defined by the loans made by international organizations to finance educational initiatives in Brazil.

Another problem with the hiring of third parties for testing and analysis which our research identified has to do with the contracts which technicians hold with these institutions and companies. Often, these pay salaries that are greater even than those of the directors of the state educational systems which are to be assessed. Because of this, when the contracts are finished, none of these technicians has any desire at all to join the state assessment teams, as this would result in a substantial reduction in pay. We even identified a case in which one of these technicians opened their own company in order to directly offer services to the state educational system.

Finally, given the technical complexity of conducting and analyzing large scale assessments, the number of institutions able to compete in the contracting process is very small. The field can, in fact, be considered an oligopoly. Confronted with this problem, one of the states under study even hired a foreign company to conduct their assessment procedures.

ASSOCIATION OF RESULTS WITH INCENTIVES

Although the international literature notes that a global trend exists towards associating budgetary incentives with positive assessment results, in many educational systems and in general in classic liberal thought, these initiatives are polemical and thought to increase inequality. Such measures are sometimes couched however as a way of reducing inequality, by giving cash rewards to under performing sectors that manage to improve their results. Although this option has not been explicitly adopted by the systems under study, in at least two of them, some sort of financial award has been tacitly adopted. In one system, a cash prize (increased salaries for school employees) is given out for the schools at the top of the

student performance rankings. In the second system, a composite indicator has been constructed using a series of performance variables (dropout rates, passing rates, attendance, etc.) which includes students' test results and which associates good performance with financial gains. An understanding that assessment must generate consequences has been evidenced by several of the managers that we interviewed. One initiative which has been outlined is linking students' assessment results to career benefits such as increased salaries or bonuses or even to higher levels of certification as an educator. Such initiative has been interpreted as an alternative that articulates assessment and education system management.

In some systems, managers have manifested doubts regarding the quality of their systems' teaching professionals, a fact that has consequently diminished interest in assessing them. Managers recognize, as well, the political difficulties facing such a measure. An example of the problems faced by teacher assessment can be found in the resistance encountered by the federal government when it tried to propose one of these programs in 2003. Faced with these difficulties, "indirect" initiatives have been developed to evaluate teachers' quality, such as asking them to choose and respond to the question which they consider to be the most interesting on the students' tests, or asking professors to elaborate a correction key for the exam and checking this for discrepancies against the official key.

INCREASING THE VALUE OF ASSESSMENT RESULTS

Though general criticisms are often made (even by system managers) regarding the limits of the assessment systems under analysis here, a large number of our interviewees affirm the assessment itself is quite important.

Although we cannot qualify this as the successful implantation of an "assessment culture" – an objective often referred to in the literature regarding these initiatives – such opinions indicate the sedimentation of the idea of assessment's importance, at least among some of the system managers. In the words of the World Bank's representative, mentioned in this study, assessment today plays a bigger role in the debates regarding the Brazilian educational system than it did a decade or two ago, a fact which is in and of itself indicative of the success of past assessment initiatives.

One observes, however, that the importance of assessment has not yet become linked to assessment of implemented policies and thus to assessment of the educational system as a

whole. So far, the tendency has been to emphasize assessment's use as a control mechanism for school performance, exploring consequences only in terms of what might happen to the school. It's still rare to encounter a view of assessment as something that can be applied to the formulation, implementation and results of educational policies as a whole.

In any case, assessment is substantially different when it is a component of educational policy and when it's directed at the center of the educational system, given that by its very nature – diagnosing, reflecting and acting – it encompasses a form of action and transformation that is not always present in other forms of implementing educational policy.

FINAL CONSIDERATIONS

A relationship between assessment and improvements in the quality of education still does not exist. A quality control dimension to teaching does not appear to have cohered in the cases under study, either in management, rationalization of resources, or in terms of specific policy initiatives based on assessment. Even the competition that some measures (such as the awarding of money or other prizes) seem to have stimulated as not resulted in advances in terms of new practices.

In short, the technical advances in assessment have not been met with a corresponding advance in using assessment's power to induce policies and practices. It must be recognized that educational systems have improved the procedures used in order to measure students' performances, but a corresponding improvement has not yet been observed in assessing the educational systems themselves. What little evidence that exists of a trend in this direction is still quite tenuous.

Usually, the concept of quality that has been forged has been restricted to appreciating student performance, without interpreting this in the light of intra- and extra scholastic contexts.

The assessment systems thus tend to limit themselves to information that's directly related to education. Doubtless, the production of such information is indeed a step in the assessment process, but the process itself is only fully realized when the information gathered is used to make judgments, take decisions and enter into action. In other words, assessment is ultimately only truly meaningful when it subsidizes interventions that transform and democratize education in terms of access, permanence and quality. It cannot be left essentially

to supporting repeated denunciations of low quality education, for the costs of its implementation are too high for assessment to simply repeat what we already know.

Such views are already being expressed by educational system managers and have pushed some to seek out new guidelines and systems for assessing educational systems. One of the movements that appears to be the most promising in this sense is the attempt to link external assessment with internal assessment, with the goal of not just focusing on schools but on all levels of the education system. In the opposite direction, we find initiatives which link assessment performance to financial rewards and which probably impact upon the internal organization of school work and the relationships between schools, with consequences that, as of yet, cannot be guessed. Further studies focused on these new factors and their effects in terms of inducing practices and behaviors are thus called for.

BIBLIOGRAPHIC REFERENCES

ABICALIL, C. A. “Sistema nacional de educação básica: nó da avaliação?” *Educação e Sociedade*, Campinas, v. 23, n. 80, p. 255-276, Sept. 2002.

AFONSO, A. J. *Avaliação educacional: regulação e emancipação: para uma sociologia das políticas avaliativas contemporâneas*. São Paulo: Cortez, 2000.

_____. “Reforma do Estado e políticas públicas: entre a crise do Estado-nação e a emergência da regulação supranacional”. *Educação e Sociedade*, Campinas, n. 75, p.15-32, 2001.

ANDRADE, D.; KLEIN, R. “Métodos estatísticos para avaliação educacional: teoria da resposta ao item”. *Boletim da ABE*, v.15, n.43, Sept/Dec.1999.

ARELARO, L. R. G. *A (Ex)tensão do ensino básico no Brasil: o avesso de um direito democrático: uma análise da ação governamental nos últimos 25 anos (1962-1987)*. 1988. Doctoral thesis in education – Faculdade de Educação, USP, São Paulo.

BAETA, A. M. B. “Pressupostos teóricos e implicações práticas do Modelo Saeb-2001”. *Estudos em Avaliação Educacional*, São Paulo, n. 25, p.23-42, Jan./Jun.2002.

- BARRETTO, E. S. S. A “Avaliação na educação básica entre dois modelos”. *Educação e Sociedade*, Campinas, v. 22, n. 75, p. 48-66, Aug. 2001.
- BARRETTO, E. S. S.; PINTO, R. P. (Coord.). *Avaliação na educação básica:1990-1998*. Brasília: MEC/Inep/Comped, 2001. (Série Estado do Conhecimento, n. 4)
- BONAMINO, A. C. *Tempos de avaliação educacional: o Saeb, seus agentes, referências e tendências*. Rio de Janeiro: Quartet, 2002.
- BONAMINO, A.; FRANCO, C. Avaliação e política educacional: o processo de institucionalização do Saeb. *Cadernos de Pesquisa*, São Paulo, n.108, p.101-132, Nov.1999.
- BONAMINO, A.; BESSA, N.; FRANCO, C. (Org.). *Avaliação da educação básica: pesquisa e gestão*. Rio de Janeiro: PUC-Rio, Loyola, 2004.
- BAHIA (Estado). *Avaliação de desempenho do ensino fundamental do Estado da Bahia 2001*. Salvador: SEC, UFBA – SP – Fapex, 2003. (Educar para vencer: projeto de avaliação externa)
- BRASIL. *Constituição da República Federativa do Brasil*. Brasília: Senado Federal, 1998.
- BRASIL. Ministério da Educação. Inep. *PISA 2000: relatório nacional*. Brasília, 2001.
- BROADFOOT, P. *Education, assessment and society: a sociological analysis*. Buckingham: Open University, 1996.
- BROOKE, N. “O Futuro das políticas de responsabilização educacional no Brasil”. *Cadernos de Pesquisa*, São Paulo, v.36, n.128, p.377-402, May/Aug. 2006.
- CANO, I. *Introdução à avaliação de programas sociais*. Rio de Janeiro: Editora FGV, 2002.
- DARLING-HAMMOND, L. “The Implications of testing for quality and equality”. *Phi Delta Kappan*, Bloomington, v.73, n.3, p.220-225, Nov. 1991.
- DIAS SOBRINHO, J. “Campo e caminhos da avaliação: a avaliação da educação superior no Brasil”. In: FREITAS, L. C. (Org.) *Avaliação: construindo o campo de a crítica*. Florianópolis: Insular, 2002. p.13-62.

FRANCO, C. “Quais as contribuições da avaliação para as políticas educacionais?” In: FRANCO, C.; BONAMINO, A.; BESSA, N. (Org.). *Avaliação da educação básica: pesquisa e gestão*. Rio de Janeiro: PUC-Rio; São Paulo: Loyola, 2004. p. 45-63.

_____. O Saeb – “Sistema de Avaliação da Educação Básica: potencialidades, problemas e desafios”. *Revista Brasileira de Educação*, São Paulo, n. 17, p.127-133, May/Aug. 2001.

FREITAS, L. C. *Avaliação de sistemas públicos de ensino: construindo metodologias alternativas*. 2002. Mimeo.

HEUBERT, J. P.; HAUSER, R. M. (Ed.). *High stakes: testing for tracking, promotion and graduation*. Washington: National Academy, 1999.

IAIES, G. Evaluar las evaluaciones. In: TEDESCO, J. C. (Org.). *Evaluar las evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa*. Buenos Aires: Unesco/Iipe, 2003. p.15-36.

KAWAUCHI, M. *Saresp e ensino de história: algumas questões*. 2001. Master's dissertation – Faculdade de Educação, USP, São Paulo.

KOHN, A. *The Case against standardized testing: raising the scores, ruining the schools*. Portsmouth: Heinemann, 2000.

LANDSHEERE, G. *A Pilotagem dos sistemas de educação: como garantir a qualidade da educação?* Porto: Asa, 1997.

LEVIN, B. *Reforming education: from origins to outcomes*. London: RoutledgeFalmer, 2001.

MACHADO, C. *Avaliação e política – limites e potencialidades: uma análise do Saresp*. 2003. Doctoral thesis in education – Faculdade de Educação, USP, São Paulo.

MADAUS, G. F. “The Influence of testing on the curriculum”. In: TANNER (Ed.). *Critical issues in curriculum*. Chicago: University of Chicago, 1988.

MCNEIL, L. *Contradictions of school reform: educational costs of standardized testing*. New York: Routledge, 2000.

OECD. *Knowledge and skills for life: first results from Pisa 2000*. Paris, 2001.

_____. *Measuring student knowledge and skills: the Pisa 2000 assessment of reading, mathematical and scientific literacy*. Paris, 2001a.

OLIVEIRA, R. P. “Avaliações externas podem auxiliar o trabalho pedagógico da escola? In: São Paulo”. Secretaria Municipal de Educação (Org.). *Educação: fazer e aprender na cidade de São Paulo*, 1. São Paulo: Fundação Padre Anchieta, 2008. p. 230-237.

OLIVEIRA, R. P. “Qualidade com garantia de respeito à diversidades e necessidades de aprendizagem”. In: GRACIANO, M. *O Plano de desenvolvimento da educação (PDE)*. São Paulo: Ação Educativa, 2007. p.32-34. (Questão, v.4)

OLIVEIRA, R. P.; SOUSA, S. M. Z. L. “Avaliação de curso: uma dimensão da avaliação na universidade”. *Revista Adusp*, São Paulo, n.18, p.30-35, Oct. 1999.

ORFIELD, G.; KORNHABER, M. L. (Ed.). *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: Century Foundation, 2001.

PELEGRINI, S. M. *Formação de professores: avaliando conteúdo e habilidade em língua portuguesa*. 2000. Masters dissertation – Universidade do Oeste Paulista. Presidente Prudente.

PERRENOUD, P. “A Avaliação dos estabelecimentos escolares: um novo avatar da ilusão cientificista?” *Ideias*, São Paulo, n.30, p.193-204, 1998.

ROGGERO, P. “Avaliação dos sistemas educativos nos países da União Europeia: de uma necessidade problemática a uma prática complexa desejável”. *Eccos Revista Científica*, São Paulo, v. 4, n.2, p.31-46, Dec.2002.

SÃO PAULO (Estado). Secretaria da Educação. *Saresp 96/97: resultados comparativos*. São Paulo: SEE/FDE, 1998.

_____. *Saresp 97/98: resultados comparativos*. São Paulo: SEE/FDE, 1999.

_____. *Diretrizes educacionais para o Estado de São Paulo, no período de janeiro de 1995 a dezembro de 1998*. São Paulo: SEE, 1995. (Comunicado)

SILVA, I. L. F. *Reforma ou contrarreforma no sistema de ensino do Paraná: uma análise da meta da igualdade social nas políticas educacionais dos anos 90*. 1998. Masters dissertation – Faculdade de Educação, USP, São Paulo.

SOUSA, S. M. Z. L. “Avaliação do rendimento escolar como instrumento de gestão educacional”. In: OLIVEIRA, D. A. (Org.). *Gestão democrática da educação: desafios contemporâneos*. Petrópolis: Vozes, 1997. p.264-283.

_____. “Avaliação e políticas educacionais: iniciativas em curso no Brasil”. In: HIDALGO, A. M.; ILEIZI, L. F. S. (Org.). *Educação e estado: as mudanças nos sistemas de ensino do Brasil e Paraná da década de 90*. Londrina: UEL, 2001. p.69-98.

_____. “Possíveis impactos da avaliação externa no currículo escolar”. In: ROSA, D. E. G.; SOUZA, V. C. *Políticas organizativas e curriculares, educação inclusiva e formação de professores*. Rio de Janeiro: DP&A, 2002. p.23-38.

SOUSA, S. M. Z. L.; ALAVARSE, O. M. A “Avaliação nos ciclos: a centralidade da avaliação”. In: FREITAS, L.C.; GATTI, B. A.; SOUSA, S. M. Z. L. (Org.). *Questões de avaliação educacional*. Campinas: Komedi, 2003. p. 71-96. (Série Avaliação: construindo o campo e a crítica).

SOUSA, S. M. Z. L.; OLIVEIRA, R. P. “Políticas de avaliação da educação e quase mercado no Brasil”. *Educação e Sociedade*, Campinas, v. 24, n. 84, p.873-895, Sept.2003.

SOUSA, S. M. Z. L. et al. “Avaliação educacional e gestão: iniciativas no âmbito do Município e do Estado de São Paulo”. In: REUNIÃO ANUAL DA ANPED, 23: GT5: Estado e Política Educacional: Política, Gestão e Financiamento da Educação, 2000, Caxambu. *Anais...* São Paulo: Anped, 2000. p. 429-435.

TORRES, R. M. *Que (e como) é necessário aprender? Necessidades básicas de aprendizagem e conteúdos curriculares*. 2. ed. Campinas: Papirus, 1995.

VIANNA, H M. “Implantação de avaliação de sistemas educacionais: questões metodológicas”. In: CONHOLATO, M. C. (Org.) *Idéias 30: sistemas de avaliação educacional*. São Paulo: FDE, 1998. p. 147-160.

VIANNA, H. M. “Medida da qualidade em educação: apresentação de um modelo”. *Estudos em Avaliação Educacional*, São Paulo, n.2, p.99-104, Jul/Dec. 1990.

WILDE, S. *Testing and standards: a brief encyclopedia*. Portsmouth: Heinemann, 2002.

Received: February 2010

Approved for publication: June 2010