

# Critical notes on standardized evaluation: loss of quality and social segmentation

JUAN CASASSUS

jasassus@vtr.net

Metropolitan University of Educational Sciences, Chile

## ABSTRACT:

The quality of education and its conceptual formulations and practices with regard to evaluation are the major concerns of the author of this paper. Its objective is to reveal how normative standardized tests, like those of centralized evaluation systems, do not serve their socially-intended purposes, since they give no information on educational quality. The author draws attention to the excessive power given to standardized assessment instruments in the evaluation and management of educational systems, arguing that, contrary to their stated objectives, they promote inequality and reduce the quality of education.

## KEYWORDS:

Evaluation and the quality of education, Standardized assessment, Educational management, Social segmentation, Loss of quality.

## INTRODUCTION

In the early 1980s, there occurred a revolution in world education policy whose reverberations soon reached Latin America. With this revolution, the focus of educational policy shifted from the expansion of the system itself to what was occurring within the already existing system. At the time, this revolution was described as a move away from a quantity-centred focus and towards a quality-centred focus. The objective of the new educational policy, therefore, was to change the way the system was regulated, based on the then-nascent discipline of educational management. Sadly, the hopes that many of us nurtured in those days came to nothing, and we have to acknowledge that the policy conceived back then has failed in its objectives. One key aspect of this failure was that no-one ever defined what they meant by “quality of education”. There was no debate, in fact, on what quality was, and therefore there was no attempt even to arrive at a consensus. It’s surprising that this word, “quality”, should have remained so empty of content. Just like the debate about what intelligence was, and the difficulty of defining it, of the early 20th century, with the conclusion that intelligence was what was measured in intelligence tests, the question of quality of education boiled down to a similar conclusion: quality was seen as equivalent to a grade in a standardized test. In more than one sense, intelligence tests were the precursors of the assessment systems used for determining the quality of education.

This absence of content and the absence of consensus on what quality meant have proved extremely problematic. For reasons yet to be explained, it was taken for granted that everyone knew what quality of education was all about — to an extent that since those days, it has been enough to justify educational policy by saying it was to “improve the quality of education” to have it accepted by the authorities. However, the policy in question was *management*, not *educational*, policy. Educational policy has since then centred on the management of the education system, and not on education itself.

Just to illustrate the topic, let’s look at some policies designed to improve education: more classroom hours, longer school years, decentralization, assessment (lots of standardized, centralized assessment), a focus on directors and their management, competition between schools, privatization etc. But none of these policies so beloved of national and international authorities had anything to do with education. What they had to do with was the management of the education system. Their origins did not lie in the disciplines which address education and its objectives. Not philosophy, nor psychology, nor sociology. Their origins lay in economics.

If we look at the criteria they established, we can affirm that these policies, which have been implemented for the last 25 years, have been a failure: not only do the results produced by systems of evaluation fail to show any progress, but many countries have actually recorded a deterioration (i.e. grades

have fallen), accompanied by an exacerbation of social inequality and segmentation.

Educational policy is a difficult matter, for education is not an exact science. It requires caution and common sense, but it also requires willingness to change. The time to change has come. It's time to apply some common sense and set out the problem in terms of some basic proposals:

1. The purpose of education (what I mean by quality of education) is to make people better, and thereby to make society better. This requires in-depth reflection on the purposes of education in the 21st century — what type of people do we want to produce, what type of society do we want to produce? And this means putting an end to simplistic thinking about education which evades the real debate and restricts itself to the wielding of instruments.
2. Put education at the service of children, not at the service of authority. What this means is that instead of an education system whose purpose is subjecting students to the moral judgement of grown-ups, we try to create an education system which gives students the experience of respect and living in harmony with others, which makes them more independent, which develops their ability to think and reflect, which awakens their feelings of empathy and solidarity, and which above all does not strip them of their powers and personal dignity.
3. Work with teachers, not against them. This means it's necessary to provide teachers with assistance in their work, listening to them, trusting them, asking them for advice on how to make things better, giving them time, and above all refraining from applying constant pressure, whose only effect is to depress and sicken them.
4. Bear in mind that it is processes that shape the terms of engagement which people cultivate with themselves and with others. Therefore it is extremely important to work at the level of the processes which are active in schools (interactions and climates), instead of assuming that educational work is all a question of measuring results.

The objective of these notes is to conduct a critical appraisal of current policy and to articulate the

reasons that lead us to think the way we do. These reasons are based on the empirical evidence that current policy has failed, that the emphasis on one type of evaluation — assessment — in particular leaves the system trapped in a pernicious dynamic regarding its objective of improving quality, a consequence of which is the opposite of the desired effect, i.e. a reduction in quality, which is impoverished in its scope and no more than an instrument of management which robs teachers of the will to improve education. Centralized assessment systems empty education systems of the purposes for which they were created.

## HAVE THE PREVALENT EDUCATIONAL POLICIES FAILED?

Educational policy in the nineties, and the educational reform which swept Latin America in this period (as it did other parts of the world), were part of a major effort to revitalize energies which had been sapped by the costs of structural changes in the world of education during the eighties. In the early nineties, governments raised teachers' salaries, built new schools, distributed many more books and enacted new laws. At the same time, it was proclaimed that progress in education would be reflected in an increase in scores in multiple choice tests.

Despite all the effort expended all over the world, however, not a single country which measures academic performance registered an increase in scores. Tests administered as part of international comparative studies (TIMSS, LLECE, PISA) as well as national assessments based on score evaluations delivered disappointing results. Education, in the light of these criteria, seemed to have reached stagnation point. If we accept these criteria, the empirical evidence indicates that the effort expended failed to yield the expected results: in some countries, such as Brazil, Chile, Spain, France and the USA, grades have fallen and inequalities persist. The reaction of the authorities to this failure has not been an acknowledgement that a change of policy is necessary, but instead a hardening of their current policy, with more standardization and more quantitative evaluation. This position can only make the situation worse.

In my perspective a crucial factor in the failure of education at present is its system of management, and particularly the system of evaluation and the theories which underpin it. The official evaluation system — which is the principal instrument of management — has become an obstacle to the development of education, encouraging regression and contributing to the reduction in the quality of education. Below we will briefly examine some of the arguments which corroborate this verdict.

### THE USE OF STANDARDIZED TESTS AND THEIR LIMITATIONS AS TOOLS FOR THE EVALUATION OF THE QUALITY OF EDUCATION

The multiple choice tests used in national evaluation systems such as the SAEB in Brazil and the SIMCE in Chile are known as “standard-indexed tests”. The objective of these tests is to determine positions and establish rankings among the individuals who take them. These rankings are constructed relative to a standard, constant average of 250 points. The creator of these tests, Robert Glaser, was careful to point out that they measure neither what students know nor what they can do. To evaluate what students know and can do, we need another type of assessment, such as portfolios, performance tests or constructivist approaches. The only ability which these standard-indexed tests measure is, according to Glaser (Glaser, 1963, 1997), the ability to remember procedures (in mathematics) or to recognize the right answer from a number of alternatives. It is a conceptual error to affirm that this type of test measures what school students know or can do. And it is an even worse error to equate the grade obtained with the quality of education given. Grades do not equal quality.

Quality is not expressed by grades. A quality education is what we might call a “good” education. A quality education essentially has to do with the ability of the school to help students become better people, so that society becomes a better place. It’s a process of transformative knowledge. The level of quality of the education provided by a school is proportional to the depth of analysis placed at the disposal of the students, the types of questions they

are allowed to ask, the types of project they can get involved in and the types of problem they are capable of solving. Educational quality does not boil down to higher grades. Quality and grades belong to different spheres, and it is an error to mistake one for the other, for the consequences are negative.

### HOW EXCESSIVE EMPHASIS ON THE INSTRUMENTS OF ASSESSMENT IN EVALUATION NEGATIVELY AFFECTS THE QUALITY OF EDUCATION

Voluntarily or not, this type of evaluation has been legitimized and empowered as the chief vector of educational policy. The results of evaluation tests form the basis for policies of accountability, incentives and penalties, the focus on resources, orientation of abilities, use of information for decision-making etc. In the light of the stated intentions to improve the quality of education, this is a contradiction. If we analyse what happened with the implementation of these systems, we’ll see that in practice they are a mechanism not for improving the quality of education but for making it worse. It’s worth noting in passing that it is a political error to assume that the success or failure of a policy depends on higher grades in an assessment which does not measure what it is supposed to, and that the system as it is designed is unable to change.

We might ask ourselves, *why has this type of evaluation, as part of a management model, become an instrument which reduces quality when it’s actually supposed to increase the quality of education?* The answer is that not only has it become an obstacle to the development of education, it is actually — contrary to what is proclaimed in the media — causing the system to regress. This assertion can be corroborated from various angles. By way of example we can cite:

1. Falling grades. As indicated above, the most cursory examination of comparable statistics shows that, despite all the efforts, grades continue to fall. Faced with this problem, the authorities easily fall prey to the temptation of intensifying their policy by applying more pressure on teachers, instead of reviewing the system.

2. Perverse effects of tests. The importance attached to tests as a form of evaluation has led to “quality of education” becoming synonymous with test grades. This is a dangerous misidentification, with various harmful effects:
  - a. it shifts the focus of education on to answers to psychometric tests and away from teaching, which is the foundation of the profession,
  - b. it overlooks things which are important objectives of education, such as: the development of personality, respect, good citizenship, cultivation of values, the will to acquire knowledge, commitment to learning etc. None of these aspects is evaluated by psychometric tests. And so the essentials of education are pushed to one side in the educational process, simply because they are not evaluated.
3. Raising grades is an exercise in illusion. Equating quality of education with higher grades in evaluation testing merely generates illusions. Training students to reply to standardized tests is like reducing a fever with medication or a cold compress: we take the patient’s temperature, see that it has reduced, and conclude that the patient has recovered. Similarly, grades can be artificially inflated or distorted in many ways, for example:
  - a. Inflating scores (as occurs with university entrance exams)
  - b. Removing the “bad” students when tests are approaching (this “bad” has no specific meaning but is merely subjective)
  - c. Giving test results to students
  - d. Training students to reply to the tests
  - e. Students are at liberty to answer at random.
4. Education becomes superficial. Focusing policy on tests instils a dynamic which makes education superficial.
  - a. The ability to respond to multiple choice tests does not involve reflecting on how to reach the right answer, but merely the act of selecting one answer. Learning to think things through is important, for it is through reflection that we look into things more deeply, and thereby increase our mastery of the content.
  - b. Responding to multiple choice tests requires neither thought nor reasoning, but merely the ability to recognize the right answer. Learning to think and reason is important, for it allows us to identify logical structures and make connections between events.
- c. Doing tests of this type does not require the ability to build knowledge, but merely the ability to “point out” an answer already known.
- d. “Exam-oriented teaching” requires teachers to spend their time training students how to make the correct choice. This means less time for teachers to teach their students how to think.
- e. Raising or lowering grades is a superficial remedy which has nothing to do with quality education. To think that an educational system improves (or gets worse) because grades increase (or go down) is to think that aspirins and cold compresses make a fever disappear.
5. Social consequences: increased inequality.
  - a. In schools with limited resources, exam-oriented teaching “to increase grades ... at any cost” means teachers are forced not to *educate* but to *train* their students: in how to do multiple choice tests. In the more influential sectors of society, education is paid for outside classroom hours.
  - b. It orients expenditure towards tests, not teaching. The theory which underpins this perspective assumes that the family’s level of income determines the outcome of the test. The differences in the results of these tests are not due to educational factors. And so, as the result is known beforehand, and since variations are known to be marginal, instead of continuing with evaluation (and paying the price) it would be more useful to apply resources detailed to evaluation to educational activities which are richer and more productive for students.

## THE ERROR OF USING STANDARDIZED TESTS AS AN INSTRUMENT FOR THE MANAGEMENT OF THE EDUCATION SYSTEM

Evaluation based on standard-indexed psychometric tests is geared towards selection, not inclusion. Therefore, in segmented contexts such as those of



the education systems of Latin America in general and Chile in particular, they foment inequality. This is the “dark side” of the question, and it has two facets: assessment systems are politically charged, and the use of information is politically charged.

IQ tests and national assessment systems are conventionally described as “objective”, but in reality they are not: they are subject to the bias and the specific conditions of those who compile them. And this insistence on their “objective” character has a perverse effect: the assumption that intelligence and mental development are equally distributed across society. Which is to say, it’s pure chance that some students find themselves at the “stupid” end of the spectrum and others at the “intelligent” end. If IQ tests or national assessment systems really were objective, the assumption would be that there would be as many rich students as poor at the “stupid” end of the spectrum, and as many poor students as rich at the “intelligent” end, given the shape of the distribution curve. But this is not the case. Test results show that poor students are found at the “stupid”, low-scoring end; and rich students occupy the other extreme, the high-scoring, “intelligent” end. This is the only valid and consistent information this type of test yields: that poor, inner-city schools have bad grades because it is “scientifically” indicated that this is where the “stupid” students are to be found. And they just happen to be poor. A merit-based system in a formal democracy, i.e. a democracy which does not take inequalities of conditions and contexts of poverty into account, tells us that the stupid deserve to be poor because they are stupid. This is to overlook a crucial point: the negative effects of poverty on the learning processes and cognitive development of students.

Assessment systems in which information circulates freely, in a formal democracy such as Chile’s, on the assumption that it allows students to situate themselves in relation to their generation, do precisely this — place them at the top or the bottom of the achievement curve. The political reason for this is that the system in place is a structure of social control and the use of this type of information is the mechanism of perpetuating social control. This leads the needier sectors of society to a conviction that their low grades are because they are stupid. It “makes sense”, therefore, that they are where they

are, and it also “makes sense” that the rich are where *they* are. It’s “right” therefore, that the economic elite should be in charge, for they are the intelligent ones, and that the poor should do as they’re told, for they are the stupid ones. The ultimate implication is that things are the way they are for reasons of eugenics.

Furthermore, it’s an illusion to think that this system serves to improve the quality of education. It does not enable teachers to give better classes, and neither is it useful for management. It generates the illusion that we are taking the pulse of the system (see the analogy of the fever and thermometer above) but in fact it is not even rooted in reality. To use it as an instrument of management only serves to identify positions and rankings, or to put this in other terms, to segregate and fragment. This has the consequence of causing much unnecessary and unjust suffering at every level of the system.

This paper does not address the question of evaluation in general — and some kind of evaluation is necessary — but a particular kind of evaluation whose consequence is not improved quality but increased pressure on teachers in the form of a plethora of incentives and penalties. It’s difficult to understand the logic of the underlying theory, according to which teachers will teach better if they are punished (or rewarded). The literature on management and the managerial practices of ministries and offices of education amply demonstrates that this system does not work. It may occasionally produce some short-term gains, but these rapidly disappear.

It’s normally assumed that rewards and penalties are effective incentives to action and are easy to implement. Note, however, that when we say “incentives” we are talking of an external (extrinsic) motivation, a carrot-and-stick brandished by the system to make the teachers work. But reservations have recently been voiced on whether extrinsic motivation is an effective method for the authorities to get what they want out of the system. It does, however, work for other purposes. Extrinsic motivation as a political instrument wielded by the state is based on false premises and has perverse effects.

1. It diminishes the dignity of the teacher. The premise is that teachers have no intrinsic motivation to do their work. This is a false premise and it has counter-productive effects. We have

to ask ourselves: why assume that teachers have no intrinsic motivation to do their job well? Assuming that they lack this motivation, as a state policy, is to strip the dignity from the teacher and his work. This loss of dignity is not offset by a higher salary.

2. It undermines intrinsic motivation. There are other perverse effects too, such as the systematic assumption that teachers are not motivated to do their work, which systematically undermines their intrinsic motivation. A good education depends on the intrinsic motivation of teachers. A policy which denies intrinsic motivation is a policy which encourages education of inferior quality.
3. It makes processes more rigid. Imposing standards and evaluations not only encourages extrinsic motivation, it makes procedures more rigid and standards more standardized. On the one hand, this suits teachers who do not feel motivated to respond creatively to the problems that teaching poses. And on the other hand, it inhibits the teachers who *do* feel motivated: it “clips their wings” in their endeavours to find the most suitable ways of teaching their students, especially in the case of a constructivist curricular policy.
4. It destroys the classroom environment. It creates enmity between teachers and students and weakens trust and ties between them. The selection system means that “bad” students are sent packing to inner-city / municipal schools, which are obliged to receive them. In a healthy, convivial system, a student with difficulties is the challenge and the *raison d’être* of the work of teachers in poorer schools. But in a system which operates on the basis of grade-related judgements, where the survival of the teacher depends on the grades obtained by his students, a “bad” student can only be seen as a threat to the teacher, a problem which the teacher would rather did not exist.

## CONCLUSIONS

One thing we have to ask before we end these notes is how things have come to this state. What happened for us to abandon the tradition of educating the human being, showing concern in our students, educating to improve society? Why have teachers been forced to relinquish their enthusiasm for their job, forced into playing the game of pressures and threats? These are questions the author has been asking himself, and the painful search for answers has led him to conclude that all this has happened because education — and more particularly evaluation — has become a very lucrative business. As long as we continue to think that scores obtained in psychometric tests are an indication of quality, the authorities will be prepared to buy testing systems. And they’ll do so with the enthusiasm of one who believes himself to be doing the underprivileged a favour. They’ll do so in the belief that to train the underprivileged in a system of instructions on how to do tests which cost a lot of money to buy is to offer them a quality education... and they won’t see, as we saw above, that what they’re really offering is an impoverished education, for the test-and-standards paradigm concentrates precisely on those aspects, such as the memorization of information, which the tests are capable of measuring. Standardized examination systems and psychometric tests leave out all that is most difficult to measure: learning to think, learning to respect, learning to live with others, learning to ask relevant questions and find the answers to them, to look for the evidence of knowledge, to determine what’s important and valuable, to learn from the context. In other words, everything that’s at the heart of a quality education.

## BIBLIOGRAPHICAL REFERENCES

GLASER, R. (1963). Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist*, 18, pp. 519-521.

GLASER, R. (1997). Education for all: Access to learning, and achieving usable knowledge. *International Seminar on Educational Evaluation*.

Ministry of Education, Rio de Janeiro, November 30-December 3.

Translated by Mark Ayton