ΣΟΦΙΑ–SOPHIA

Mass measurements: A political production of meanings on educational systems

Carmenza Sánchez Rodríguez

Universidad Santo Tomás, Universidad Pedagógica Nacional de Colombia. casaro94@gmail.com

How cite

Sánchez, C. (2017) Mass measurements: A political production of meanings on educational systems. *Sophia,* 13(1) 64-74.

ISSN (electrónico): 2346-0806 ISSN (impreso): 1794-8932

UNIVERSIDAD La Gran Colombia
Fundada en 1951

Sophia-Education, volumen 13 Isuue 1. English versión

Abstract

This article includes reflections resulting from documentary reviews and discussions with teachers and experts in educational evaluation. This has allowed us to transcend the understanding of evaluation from a technical practice to a political practice produced from a world view that legitimates and installs certain meanings on the educational systems and their actors, generating specific forms of relationship and organization that produce exclusion. In this perspective, we analyze some of the meanings produced by mass measurement practices that, presented as a technical practice, obviate the discussion of the worldviews, the conceptions and evaluations from which they are produced. There are discussed concepts about evaluation, the assessed knowledge, recognized cultures and the processes of exclusion generated by these practices, as well as their effects on institutions and persons. It is concluded that tests are inventions of the twentieth century, and that the pedagogical and technical problems (that) they conceal have not been discussed; but especially, there have been ignored their implications in terms of quality and equity and their political role in the reproduction and legitimation of differences.

Keywords: Equity, evaluation, exclusion, quality, massive tests.

"Tests are twentieth century inventions, and the pedagogical and technical problems that they hide have not been discussed, but there have been especially ignored their implications in terms of quality and equity and their political role in reproducing and legitimating differences."

Carmenza Sánchez Rodríguez (2016)

This article understands evaluation[1] as a political practice, that is, as that which is produced from a worldview that installs and legitimizes certain meanings about a phenomenon; in this case, massive measurements, and the forms of relationship and organization that it creates. This practice, which in principle is understood as a technical activity, away from any subjective factor, due to its alleged neutral and objective nature, has built and has been built from certain worldviews that have legitimized conceptions about knowledge, its place of production, the role of schools in the transmission or construction of the same and the success or failure of educational systems and subjects, reproducing and, above all, legitimizing homogenization and exclusion processes.

The meanings produced by evaluation

From this conception of evaluation as a political practice, there will be performed an analysis of the meanings produced by mass measurements in the educational field.

Evaluation as neutral and objective practice

Although evaluation in the educational field has been resignified[2] as a pedagogical practice at the service of teaching - learning processes, when talking about mass measurements, evaluation is assumed as a technical practice, more associated with psychometrics than with pedagogy, which does not seem to imply positions or visions about educational processes. Not only because evaluation of quality is assumed as a technical process but because it does not call, among its criteria, the formative bets of institutions.

These measurements are made from technical approaches that aim to control biases in the elaboration of the questions, ambiguities and all the subjective or ideological components that may affect measurement. Psychometrics has tried to cover evaluation with the characteristics of "neutrality and objectivity;" for this, it has designed sophisticated statistical methods and measurement instruments that seek a strict control and monitoring of the process, thus guaranteeing the supposed equality of conditions for those who are evaluated, and "avoiding" that evaluation

be crossed by prejudices, biases, ideological or subjective elements. In this sense, both the process of preparing the questions and the application and reading of results respond to established and validated protocols to guarantee the reliability and validity of the instruments and measurements.

However, criticism of positivism and new epistemological currents have made it clear that it is impossible to guarantee the neutrality of the knowing subject, since they not only deal with the knowledge of objects or phenomena from their conceptions, prejudices, interests and visions, but also they build them from themselves. This epistemological finding not only applies to the production of knowledge, but it can be predicated of any knowledge production practice. "In this way, any evaluation is implicitly guided by certain values, although the results are always presented as 'objective' and 'scientific'" (Roth, 2010, page 6, cited by Sánchez, 2015).

The conception of evaluation as a neutral - objective practice has led, in the practice of mass measurements, that subjects assume the results of evaluation as universally valid and irrefutable judgments about the knowledge of the subjects, the quality of institutions and educational systems, without questioning the content of such evaluation, the values, worldviews and conceptions that are mobilized through it.

The conception of knowledge as universally valid or socially and historically recognized

Conceiving evaluation as a neutral and objective practice implies an understanding of knowledge according to which there are some universally valid knowledge or skills that all students must possess when they graduate from a certain (school) degree:

Massive tests, start from the postulate that there is a single knowledge, which is addressed in the same way from any context or culture. Physics, mathematics, language, science are the same in London as in Chocó; on this basis there are applied international tests of mathematics, TIMSS; and the language and math exams of the UNESCO's Latin American Laboratory for Educational Quality Assessment (Sánchez, 2015: 18).

Others affirm in a different but equally hegemonic conception, that they are socially and historically recognized knowledge that must be learned by students and taught in school. These two conceptions about knowledge, mobilized through evaluation, maintain hegemonic positions on what should be learned and taught in school and ignore the knowledge produced in non-hegemonic cultures or those produced in hegemonic cultures but which are not hegemonic knowledge. In this sense, the socially and historically recognized became such because it was selected by a dominant social group that recognized it, which considered it important to be learned in school and evaluated in the tests. In this selection process, only some knowledge was considered relevant, and that selection was made from specific world views that have made bets for specific cultures and production modes.

Massive measurements work on a limited conception of knowledge that only recognizes the hegemonic knowledge that has become, in the words of Kuhn (1962), "normal science". However, what has enabled scientific development has been critical positions against theories, the questioning of socially and historically recognized knowledge, and the questioning of scientific postulates that have become axioms[3].

The culture that recognizes and legitimizes evaluation

In the perspective of hegemonic knowledge, evaluation for a subject that does not belong to the dominant culture does not refer to the knowledge of its culture. This measurement not only ignores their knowledge, but also ignores their culture, because what is recognized and legitimized is Western, white, masculine and urban culture. The study carried out by Hederich & Camargo (2000) on cognitive style and educational achievement shows that in mass measurements, these population groups obtain the best results, that is to say, it is the culture of these that is legitimized in the evaluation processes.

However, evaluation in Latin America, given the ethnic and cultural diversity and economic conditions of its countries, has become a practice of exclusion for minorities[4]. For example, the ancestral populations (more than 90 groups in Colombia) obtain the lowest results in language tests, which evaluate communicative skill. However, some of these groups speak between 5 and 9 dialects in their community and their tradition is oral. On the other hand, the test is done in Spanish, and in a written form, which is another element indicative of the dominant culture.

The reading of the results and the "repeated failures" of these groups are read: "Not as cultural differences, but as intellectual differences" (Torrado, 2002: 33), blaming the subjects for the low results without questioning the technical limitations of evaluation, which prevents these groups from demonstrating, in real actions, the communicative skill they possess in their language. It is necessary to insist that communicative skill is related to the appropriation of the grammar of a specific language in its contexts of action and not with an abstract skill that makes a subject perform effectively in any language, although this seems to be the assumption of evaluation in this field.

The objects of evaluation: Measurable and verifiable knowledge

The processes of standardization implicit in mass measurements suppose a reduction of knowledge and formative processes, since in them it is taken into account only what can be easily measured and verified. The technical limitations of measurements mean that in a test can be evaluated only the knowledge that can be reduced to a closed response option. This implies that issues such as skills (which by definition have a creative character, because they suppose an action in a specific context; and this action is determined by the singular and specific characteristics of it), are impossible to measure in a massive test.

Tests could account for an interpretive skill, related to reading comprehension, from which could not even be inferred a communicative skill, it is difficult to argue that the tests can assess skills in other areas, since evaluation of skills is an evaluation focused on the performance that it supposes, among other characteristics: "The ability to transfer knowledge and understanding to action ... and to build and provide original answers" (Díaz, 2005: 128).

From the above it is inferred that if the quality of education is related to the ability of the school to develop skills[5], it is not possible to know the competence level of the students, since the tests are not performance evaluations.

Conceptions about teaching and learning promoted by evaluation

Massive measurements also imply conceptions about teaching and learning according to which everyone learns the same and in the same way, and therefore can be evaluated through the same test. There are also implicit conceptions of homogeneity regarding the (evaluated) subjects, which don't recognize the differences in cognitive styles, rhythms and learning styles, just to mention a few.

For example, mass measurements favor some cognitive styles over others. A study carried out by Hederich & Camargo in 2000 on cognitive style and educational achievement shows: "That the educational system favors the achievement of learning (skills) of only one of the cognitive styles considered: the least common in the population…"(p.139), which is the independent cognitive style.

These conceptions of learning, teaching and evaluation as homogeneous processes have led to standardization and homogenization of processes and subjects, ignoring a long struggle given by the social sciences for the recognition of the uniqueness of the subjects and of processes, which is not only a pedagogical issue but a political and ethical one related to the recognition of diversity and differences.

From this approach, the evaluative practice, known as a phase of the teaching - learning processes, was converted from the mean to the end of these processes. It is taught to evaluate, and it is learned to be evaluated; and the curriculum, understood not only as the curriculum, but as the relationships between the subjects (teachers-students, parents-students, teachers-managers), the relations of the school with society and knowledge, formative and pedagogical bets and the interactions among subjects, ended up reduced by evaluation. The curriculum was reduced to a study plan, and this to what is evaluated in the tests.

A study carried out by the National Observatory of Evaluation Policies (ONPE, for its initials in Spanish) of the *Universidad Pedagógica* of Colombia (Suárez 2007) shows that the institutions have reduced their study plans to what is evaluated in the tests, that the areas that are not subject to massive measurements have been relegated to a second place, which have reduced the hours for teaching them or have disappeared from the study plans. It also concludes that the Institutional Educational Projects, which in Colombia allowed the institutions to make formative bets according to the needs of their contexts, were forgotten as projects, when it became evident that the evaluations did not take into account the singularities of the institutions.

The conception about the quality of education

The conception of quality related to the results of the tests circulates in what has been called a request of principle[6]: "The quality of education is bad because the results are bad; and the results are bad because the quality of education is bad" (Sánchez, 2012: 20).

This fallacy of quality discourses compels us to question the notion of quality, since, as the ONPE has stated: "Quality is an abstract category susceptible of multiple contents" (Sánchez, 2012: 19). In the public educational policy, quality is understood from the results in the tests, that is, quality is conceived as a product. In the pedagogical field and in terms of the social function of the school, which understands education as a right, the quality of education is related to providing the necessary conditions for students to achieve their learning.

Therefore, from the perspective of equity and the recognition of differences, the school must provide the necessary tools so that students can successfully develop any life project (ranging from being a farmer, singer or sportsman to more traditional liberal professionals, such as medicine or law). The school was not created only for students who wish to continue their academic life in the universities, the school serves diverse populations that enter (to) it, with different needs and expectations, and therefore require different and differentiated learning. The ability to respond to these needs and interests must be taken into account when evaluating the quality of education.

In Colombia: "Only 20 out of 100 baccalaureates access higher education" (Caracol, 2005); this means that, either because of the economic conditions that prevent all students who graduate from secondary education from entering higher education or because not all students wish to continue higher (college) studies, 80% of the population has a life project different from the academic one. On the other hand, a country requires subjects dedicated to diverse occupations for the construction of a collective project. That some of them are not socially recognized implies that the States and the school must work to ensure that the conditions of that occupation be dignified and allow those who exercise them mobility and social appreciation, but never suppose that those who wish to dedicate themselves to them do not have a project of legitimate life, to which the school does not have the obligation to contribute. When the quality of education is understood in terms of results for academic life, legitimate and necessary life projects are unknown, the function of the school is reduced and the States are disclaimed of their obligations in generating conditions for guaranteeing quality educational processes, relevant to the expectations and interests of the students and the needs of the contexts. This without entering into the discussion about academic learning, learning for life and work performance, which rarely coincide or are articulated in the school.

On the other hand, when quality is read as a result and not as a condition for... the subjects are held responsible for the results with arguments such as skills deficiency or cognitive or learning problems, without making an evaluation of the conditions in which these subjects learn and they are evaluated.

In this sense it is worth analyzing some of the conditions in which Colombian students learn, and in general Latin Americans, who are considered essential factors to achieve learning. First, the teacher is considered one of the essential factors to ensure the quality of learning. However, in Colombia there are frequent news such as the following: "36 rectors of different school institutions ... sent an office to the mayor ... in which, in addition to the appointment of teachers, they demand the assignment of administrative personnel, counselors and the payment of public utilities" (El Tiempo, 2008); "another 250 teachers in the municipality... declared an indefinite strike because they have not been paid for four months" (El Tiempo, 1994); "Children from rural school of Puerto Boyacá have been 5 months without attending classes" (Noticias Caracol, 2015). "The **INEM** lacks both teachers and a principal" (La Nación, 2015), etc. This is the condition of many rural schools and many schools in poor and remote provinces. However, when evaluating, it is not taken into account that of the 11 years of education that are evaluated by **Saber 11** many of those years the students did not have a teacher, they only had it for some time, they hired him but never arrived; or he arrived but he had to leave because of threats from the armed groups.

Secondly, in the studies of factors associated with the results of the tests, the socio-economic conditions of the students appear as a determining factor. These studies show: "The higher the socioeconomic level of the students and their

families, the greater the expected performance in the areas and (school) grades evaluated" (Icfes, 2009). These studies show that access to cultural capital, as well as a higher educational level of the mother and the use of ICT, mean better results for students, findings that are directly linked to wealth.

Third, in figures for 2014 in Colombia, it is shown that the provinces of Chocó, Sucre and Cauca have the highest percentages of poverty and extreme poverty; they also have a high percentage of indigenous population. These data contrast with the situation in the provinces of Antioquia and Valle del Cauca, which have the lowest percentages of poverty and extreme poverty, and in which the percentage of the indigenous population with respect to the total population is little significant (UNDP, 2015). The provinces that obtained the lowest results in Saber 11[7], in 2014, were Magdalena, Bolívar, Chocó, Amazonas and Vaupés; and the cities with the lowest results were Mitú (capital city of the Vaupés province), Buenaventura (Valle del Cauca), Turbo (Valle of the Cauca), Tumaco (Nariño) and Uribía (Guajira), (El Tiempo, 2014). In these provinces and cities, there is a large concentration of ancestral and Afro-Colombian population, which as shown by the figures that are referenced and other studies, they have the highest poverty rates. The five poorest provinces of Colombia have the same peculiarity: "It is that these territories are essentially occupied by Afro-Colombians and indigenous people" (Ola Política, 2016).

Based on these results, it is concluded that the poor, Afro-Colombians, ancestral populations and women, in general and especially in mathematics[8], obtain the lowest results. For this reason, the reading that must be carried out in terms of quality assurance and equity is greater support for these populations and greater budgets to reduce gaps. However, mass measurements have assumed as a function, the selection of: "Individuals apt to continue learning" (Unesco/Orealc, 1990: 48) and the exclusion of "the less capable" (Sánchez 2006: 3).

These impoverished uses of the results are favoring not only the reproduction of the differences, but they are also legitimizing them from the discourses of the deficiencies of the subjects. The conditions of inequality are reproduced, because those who do not obtain good results cannot enter college institutions of high quality, nor to the programs that they wish, or they do not receive opportune supports to achieve their learning.

If the economic factors are separated, as they have the greatest impact on learning, teachers can significantly contribute to the achievement of learning. But if we add to the unfavorable socioeconomic context the lack of teachers, resources (library, didactic material, internet access, etc.), adequate facilities and lack of teachers training, there is little value added that the school can offer to learning.

This reproduction of the conditions that generate inequality is particularly evident in teachers[9]: "Teachers come from sectors and families with less cultural and economic capital in relative terms (Tenti, 2005, cited by Vaillant 2006) and ... in these households the incidence of economic vulnerability tends to be significantly higher than among the homes of other professionals and technicians." (Vaillant 2006).

Studies show that many of them dedicate themselves to teaching because they do not manage to enter the university or the program of their preference, due to their low results. In this sense, Denisse Vaillant (2006) states: "An educational system will not be better than the teachers it has" (p.123). However, if we reverse the reading and stop blaming the subjects (the teachers) for the results, we can affirm: "The teachers will not be better than the educational system that formed them." This circle of poverty and poor quality, generated by the poor conditions of the educational system and the unequal use and interpretation of the results, is repeated and reproduced without generating political debates on the aims and purposes of education, and on the relevance of massive measurements for assuring quality.

Is assessing (equivalent to) improving?

In the educational context, the expression "to evaluate is (equivalent) to improve" has become popular. However, after more than five decades of massive measurements, improvement has not yet been observed: "The diagnoses on the educational situation in Latin America coincide in pointing out that, in terms of quality and equity, the reality of education belies its promises." (Vaillant, 2006: 120). To evaluate is not (equivalent) to improve, nor does it imply improvement if the evaluation is not relevant and something is not done with its

results.

The uses of the results of evaluation are known, although not sufficiently discussed. It is constantly affirmed that the results serve the decision makers to generate lines of action in public policy, but this use does not come from the logic of the instruments and the measurements. When it is observed that there are no good results, standards are redefined, students are instructed in the resolution of tests, and teachers in the preparation of test questions. The accompaniment to the institutions is done on the basis of the standards, and the instruments are refined so that all the tests evaluate the same, for coherence between levels, etc. The instruments are perfected, but there are no measures that help overcome the conditions of students who do not achieve the expected learning.

The logic of the measurement and the reading of the results circulate again in a request of principle. A diagnosis is made, the diagnosis says nothing about the disease but about the symptom, the instruments are improved to measure the symptom and more and more accurate readings are made about the symptom, but neither the symptom nor the disease are treated . The perverse part of this logic is that after being sure of the presence of the symptom, because of the multiple measurements that are made[10], the subjects are excluded from the treatments because "they are not suitable to continue in the system," and it is concluded that the systems they are not of quality, without addressing the discussion about what is being understood by quality and what are their assumptions.

In the perspective of students and institutions, the results say nothing about the processes. Students who do well at school do not score well on tests and those who score well on tests do not get good results at school (Hederich & Camargo 2000). Institutions are forced to improve their results, although the results of the measurements are not comparable, and suffer the consequences of being at the lowest levels; or they receive the incentives without being able to understand why they are "good," or in what they are (good).

From the perspective of the media, institutions and subjects (students) are classified, de-contextualized readings are made about the results, and measures are promoted on schools that do not obtain good results, extolling the work of those who achieve it. From these simplistic readings of the results and the incidence of factors such as the rural, the urban, the masculine and the feminine, it is concluded that it will be necessary to change what the subjects (students) are, so that they obtain good results. The girls will have to masculinize themselves to obtain good results in mathematics, the boys will have to feminize themselves to obtain good results in citizen competitions, the inhabitants of the cities will have to ruralize themselves to improve in the citizen competence, etc. These readings not only do not promote the recognition of differences, but end up reproducing them, they also reduce comprehension and learning to the perspective of a single style, genre or population, ignoring the potential of the others.

The positive correlation shown by the studies between the results of the tests and the cognitive style independent of the environment, associated with wealth, greater access to cultural capital, closer proximity to the codes of modernity, urban and masculine inhabitants, etc. (Hederich & Camargo, 2000), requires asking if evaluation is contributing to the improvement of the quality, or to the reproduction and legitimation of the differences, and the sharpening of the processes of exclusion.

In the perspective of current epistemological and pedagogical currents, the polyphony of voices, of views, of strategies, of styles, of approaches, far from being deficiencies of a system, are factors that enrich the views, the understandings and the projects. For this reason, it is not understandable the insistence on standardization and evaluation processes that start from the assumption of the homogeneity of the subjects.

The evaluation as a mechanism to guarantee equity?

It is constantly asserted that evaluation serves to ensure that all children learn those minimums proposed in the standards ("what every student should know"), because institutions are forced to work on them to respond to massive measurements. To affirm that the tests evaluate the skills that a student must have developed at a certain age[11] or the minimum learning to a certain degree, supposes that the skills and learnings are achieved independently of the contexts. This assumption ignores the current theories about learning and

skills. The skills are developed in a context, and to affirm the opposite is to suppose that it is about innate gifts in the subjects, in which case education and evaluation would not be necessary.

For this reason, it is important to generate a pedagogical and political debate on the quality of education and equity in mass measurements. It is necessary to understand that quality is fundamentally linked to the generation of conditions so that everyone can learn, and that equity means differential attention to the needs of each subject and population. Equality in evaluation and education, in populations as diverse as ours, generates inequality and exclusion. The bet is not for an equal evaluation for all, but for an education that can provide what everyone needs.

The considerations presented are part of the constructions and relationships that have been instituted and institutionalized in the educational field, based on the results of massive tests, conceived as evaluation and referents of the quality of education.

Conclusions

Do massive tests contribute to the improvement of the quality of education?

The answer to this question, from the arguments outlined above, is a resounding NO. No, from a pedagogical point of view, for two reasons: first, because the results of the tests only serve a small number of students (the 20% who manage to enter the university); second, because the results are not used as input to support the achievement of essential learning for life (among other reasons because the tests do not evaluate the essential learning) and the formative bets do not coincide with the criteria for evaluating the tests (we hope that this coincidence does not occur because the formative aspect would be extremely reduced and simplified).

It is irrefutable that massive evaluation processes push school changes. However, as the ONPE study (Suárez, 2007) showed, these are formal and not essential, since they only give more time in the study plans to the evaluated areas, to the detriment of those that are not; they create the need of hiring training plans for the resolution of tests for the students, and the elaboration of this type of exams for the teachers, without any pedagogical reflection on its usefulness.

From the point of view of policies, the answer is equally negative because decision-makers on public policy issues (thanks to the results of the tests) have avoided discussion and research on efficient education systems, and the necessary measures to ensure that everybody learns, and that education responds to the interests, expectations and needs of students and their contexts. The discussion in public policy about the quality of education has been reduced to results and technical issues related to measurement and quality indexes (which are calculated, largely from the results), but there is no true balance of compliance of the obligation of the State to guarantee the quality of education and the supreme function of inspection and surveillance over the educational "public good," regardless of whether it is provided by private or state centers.

From a social perspective, the tests have not contributed to quality; rather, they have generated negative consequences on equity because they have allowed and fostered decontextualized readings of the results that read the cultural, cognitive, gender and learning differences as intellectual slopes that are a responsibility exclusive of the subjects (students or teachers), evading the discussions about the responsibilities of the educational systems and the governments for overcoming the conditions that generate inequity.

Evaluation has become one of the practices that generates the most inequality and exclusion, not only because it is carried out from a standardized vision of knowledge, but also because through it, there are made invisible the social and economic problems related to poor education of the poor, women and other minority groups, the low number of places available in public universities[12] that prevent students from accessing higher education, the deficiencies in teacher training (which "is believed" to be overcome by evaluating permanently; teachers are the most evaluated professionals in Latin America), the limited employment opportunity (more than 30 people undergo rigorous evaluation processes to compete for 1 position), thus contributing, not only to the reproduction, but also to the legitimization of differences .

The impossibility of the State to guarantee employment and access to the different levels of education for its citizens has been disguised with rigorous selection processes that guarantee access for the "fittest", thus generating the exclusion of the "least able," and the making of inequitable

decisions that favor those who have good results and wealth, because the correlation between wealth and results is positive and significant.

From a cost - benefit economic perspective, the balance is also negative, since the cost of measurements (not calculated so far[13]) that implies maintaining an entity dedicated to measurement, the professionals working in it, the contractors required for elaboration and review of questions, the security of the tests, the processes of printing, application, reading, interpretation, dissemination and consultation of results, and all the processes related to the divulgation of the instructions for applications, applications and dissemination of the results, among other direct costs, is not justifiable when the results are so poor and the impacts so limited in the quality of education.

From a technical point of view, the balance of the measurements also yields negative results. According to Felipe Martínez (2013), this type of tests is not designed to:

Report reliably on the most complex aspects of the curriculum, such as advanced skills in language, math, and other areas ... give precise results on the (academic) level of individual students, even at less complex points, support conclusions about the quality of individual schools or teachers by not considering the dimensions not covered by the test or other circumstances "(Martínez 2013).

The researchers and technicians of the agencies in charge of carrying out the tests have reiterated that they cannot say everything about the quality of the educational systems; they have also detected flaws in the validity of the same, in the comparability of the results, etc. However, based on the results, inferences are made about the entire education system. It is difficult to argue that a test of more or less 50 questions accounts for the training in 11 years in the area of natural sciences, which includes biology, chemistry and physics. If we start from the assumption that only what can be measured and verified is evaluated, it becomes clear that what is measurable and verifiable is the least important in the training processes, and that the quality of education is more related to students' performance in their future lives (social, academic and work), which are not possible to measure, and much less to predict through a test.

The technical problems of the tests, which have been exposed by different authors but ignored by the decision makers, raise the need to generate an informed debate about their limitations and the extent of what is said from their results. It is necessary to question the ability of the tests to account for the quality of education and the transcendental decisions that are made about the systems and the subjects (the students) based on results that say so little in relation to the students' learning and the quality of education.

The educational and pedagogical bets in the world are not looking towards standardization, much less competence-based training. However, the evaluative practice is going backwards and continues anchored in the conceptions of the psychometrics of last century. I dare to say that not only are there no pedagogical arguments to support a practice of evaluating educational systems on these assumptions and conceptions, but that this practice is not sustained, either technically or politically. Tests are twentieth century inventions, and the pedagogical and technical problems that they hide have not been discussed, but their implications in terms of quality; and equity and their political role in reproducing and legitimating differences have been ignored.

References

Ayala-García, J. (2015). Evaluación externa y calidad de la educación en Colombia. Banco de la República. Recuperado de: http://www.banrep.gov.co/docum/Lectura_finanzas/pdf/dtser_217.pdf.

Caracol (2005): Noticias. Recuperado de http://caracol.com.co/radio/2005/11/29/nacional/1133266380_225419.html).

---------- (2015): Noticias. Recuperado de http://noticias.caracoltv.com/el-periodista-soy-yo/cinco-meses-sin-clase-llevan-ninos-de-escuela-rural-de-puerto-boyaca?mobile=1#!).

Díaz Barriga, F. (2005). Enseñanza situada: Vínculo entre la escuela y la vida. México: McGraw Hill. Recuperado de: http://es.slideshare.net/nene_udelas/diaz-barriga-fridacapitulo5rubricas.

El Tiempo, Noticias. Recuperado de: http://www.eltiempo.com/multimedia/infografias/cundinamarca-y-boyaca-con-las-mejores-pruebas-saber-11/14893536

---------- Noticias. Recuperado en http://www.eltiempo.com/archivo/documento/

CMS-4098093.

Hederich, C. & Camargo A. (2000). Estilo cognitivo y logro en el sistema educativo de la ciudad de Bogotá. Bogotá: Imprenta Nacional de Colombia.

ICFES (2009) Saber 5o. y 9o. 2009. Síntesis de resultados de factores asociados Saber 11. Recuperado de: www.icfes.gov.co.

--------- (2015). Resolución 001 de la Junta Directiva. Recuperado de http://www.icfes.gov.co/index.php/transparencia/informacion-financiera-y-contable/presupuesto)

Kuhn, T. (1971). *La estructura de las revoluciones científicas*. México: Fondo de Cultura Económica.

La Nación. (2016). Noticias. Recuperado de: http://www.lanacion.com.co/index.php/noticias-regional/neiva/item/255622-el-in-em-esta-sin-rector-y-sin-profesores.

Martínez, F. (2013). Evaluación en aula y pruebas externas. Uso pedagógico e interpretación de los resultados. II Congreso Internacional de Evaluación. Equidad y Desarrollo Humano realizado el 24 y 25 de septiembre de 2013 en Bogotá. Recuperado de http://congreso.evaluacion.sedbogota.edu.co/images/memorias2013/conferencias/FELIPE_MARTINEZ%20PDF.pdf).

Noticias Caracol 2015. Recuperado de: http://noticias.caracoltv.com/el-periodista-soy-yo/cinco-meses-sin-clase-llevan-ninos-de-escuela-rural-de-puerto-boyaca?mobile=1#!

MEN. Rubros presupuestales. Recuperado en: http://www.mineducacion.gov.co/1759/articles-355956_MEN_presupuesto_2015.pdf).

Ola Política, (2016). Recuperado de: http://www.olapolitica.com/content/los-cinco-departamentos-m%C3%A1s-pobres-de-colombia

PNUD. Programa de Naciones Unidas para el Desarrollo –Pueblos indígenas. Los pueblos indígenas y los Objetivos de Desarrollo del Milenio (2015). Recuperado de: http://www.co.undp.org/content/colombia/es/home/library/mdg/pueblos-indigenas-y-los-objetivos-de-desarrollo-del-milenio/.

Roth, A. (2010). La evaluación de las acciones públicas: una actividad en constante evolución. Revista Departamento Nacional de Planeación. 1 (1). Bogotá. Recuperado de http://www.dnp.gov.co/PortalWeb/LinkClick.aspx?fileticket=OMGFW2wrRZw%3D&tabid=99

Sánchez, C. (2006). Los maestros ante la Política. Magazín Pedagógico de la Universidad Pedagógica Nacional Bogotá.

---------- (2012) Lo político, la evaluación y la calidad: Tres debates necesarios en el Observatorio Nacional de Políticas en Evaluación. *Revista Educación y ciencia*. Número 15. Año 2012, Pág. 11 – 22 Recuperado de: http://revistas.uptc.edu.co/revistas/index.php/educacion_y_ciencia/article/viewFile/2280/2858).

---------- (2005). Aproximación al Concepto de Política: Una Mirada desde el Observatorio Nacional de Políticas en Evaluación Educativa. Observatorio Nacional de Políticas en Evaluación. Boletín No. 3. Bogotá.

Suárez, Pedro. (2005). Incidencia de la actual política evaluativa en instituciones de educación básica del país. Bogotá: Fondo Editorial Universidad Pedagógica Nacional.

Torrado, María. (2000). El desarrollo de competencias: una propuesta para la educación colombiana". Bogotá: Universidad Nacional.

Unesco/Orealc. (1990). Conferencia Mundial sobre Educación para Todos. Satisfacción de las Necesidades Básicas de Aprendizaje. Santiago de Chile.

Universidad Nacional de Colombia. Noticias. Recuperado de http://www.manizales.unal.edu.co/index.php/noticias/36-ano-2014/4247-63-386-aspirantes-presentaron-pruebas-de-admision-a-la-u-n.

Vaillant, D. (2006). Atraer y retener buenos profesionales en la profesión docente: políticas en Latinoamérica. *Revista de Educación*, 340. Mayo-agosto 2006, pp. 117-140. Recuperado de: http://www.revistaeducacion.mec.es/re340/re340_05.pdf.

Footer

1 Massive measurements are not, in a strict sense, evaluation. However, in this article they are called evaluations for the purposes of the discussion, since the general public does not distinguish between evaluation (assessment) and mass testing (measurement).

2 Evaluation is not an invention of the school, but it came to it from the discourses of quality in the

business world.

3 Postulate that does not require proof.

4 Due to ethnic and cultural diversity, minority in Colombia is the generality, that is, each citizen belongs to one or more minorities (women, Afro-Colombians, displaced persons, victims, ancestral population (more than 90 peoples), LGBT, population with special educational needs, raizales, gypsies, rural, etc.)

5 In Colombia, national measurements evaluate competence (skills).

6 Circular argument in which the conclusion is one of the premises.

7 National test to evaluate the quality of Basic and Secondary education, which serves as a criterion for admission to college education.

8 "Colombia belongs to the group of countries with the lowest score in the mathematics component of the PISA test in 2012. It also recorded the largest gender gap in the score of the same test among the 65 participating countries" (Ayala- García, 2015)

9 It is not exclusive of teachers; what happens is that since they are the only professionals evaluated through massive tests in Colombia, there are results only for them, so the interpretations take place only on them.

10 In Colombia, the Saber test is applied in the 3rd, 5th, 9th and 11th (school) grades; it is evaluated to enter the university and to graduate from it, to enter the teaching career, to pass the (job) test period, annually; and to ascend in the teaching career ladder.

11This is the assumption of the PISA test.

12 To the National University of Colombia, "63,386 applicants present, in order to seek one of the 5,635 positions that the University is expected to offer" (Universidad Nacional, 2014). The University has an admission exam; that exam is passed by those who have obtained the best results in the mass measurements, since the correlation between the two tests is positive. Those who obtain better results are those who graduate from private elite schools, men, urban, high socioeconomic strata, and a few students who despite their conditions manage to enter.

13 The budget of ICFES, the institute in charge of measuring in Colombia, for 2015, according to resolution 001 of the Board of Directors was 182,614,527,380 (ICFES, 2015); while the budget for the improvement in infrastructure and endowment of institutions of basic and secondary education (regulated by Law 21 of 1982) was 155,687,900,000 pesos (MEN, 2015).