Education quality: reflections on its evaluation through standardized testing

Reflexiones en torno a la evaluación de la calidad educativa y las pruebas estandarizadas

José Eriberto Cifuentes Medina
Derly Francedy Poveda Pineda
Diego Alexander Rodríguez Ortiz

Abstract

This reflective article explores the evaluation of education quality, and whether it is necessary and relevant in the educational system, its institutions, and its processes. The risks of utilizing national and international standardized tests as the only source for testing evaluation and quality are explored, along with recognition of particularities for each context. Test results are connected to the human personnel, such as teachers, who direct and administer these exams. Thus, their efforts are considered to be reflected in these results, which do not aid in the advancement of improving quality or educational evaluation. So, the use of standardized exams becomes risky, and it can lead to systemic and systematic inefficiency.

A systematic and progressive pedagogy, along with education quality, is viable. In fact, qualitative and quantitative national and international evaluations could be used to improve gaps in education. Nonetheless, evaluation should be done as a means to improve educational processes from a general and particular standpoint that go beyond statistical indicators of logical and coherent achievement.

Keywords
Evaluation, Quality, Education, Standardized Testing

Resumen

El artículo explora la evaluación de la calidad educativa es necesaria y pertinente en el sistema educativo, las instituciones y los procesos; reconociendo el riesgo inminente de reconocer como consecuencia absoluta las pruebas estandarizadas nacionales e internacionales reconociendo los contextos particulares de cada aspecto. Los resultados están inmersos en la articulación de la calidad y la evaluación desde el recurso humano que direcciona y quien ejecuta para ver reflejados en la medición y que no avanzar en planes de mejoramiento para optimizar la calidad y la evaluación en la educación y se estaría siendo objeto de los riesgos de la evaluación de la calidad educativa en dirección de la inoperancia sistemática y sistemática.

Es viable la articulación sistemática y progresiva de pedagogía, didáctica y calidad en la educación de manera que la evaluación pueda ser un constitutivo de mejoramiento ante las falencias detectadas

DOI: 10.18041/2382-3240/saber.2019v14n2.5894

1 PhD in Education, University of Baja California - México. Master’s in Education, Specialist in Education with an emphasis in Educative Evaluation, Graduate in Theology, Graduate in Philosophy and Religious Education, Universidad Santo Tomas. Investigation group member of SIEK. University Pedagogical and Technological of Colombia. Correo electrónico: joseeriberto.cifuentes@uptc.edu.co ORCID iD: https://orcid.org/0000-0001-5702-620X
2 PhD in Education, Universidad de Baja California - México. Master’s in E-learning, Specialist in Educational Management, Graduate in Education and Information Education, Director of the School of Humanities and Education, Long-Distance Education Faculty. Investigation group member of SIEK. University Pedagogical and Technological of Colombia. Correo electrónico: derly.poveda@uptc.edu.co
3 Master’s in Pedagogy, Bachelor in Philosophy and Religious Sciences, Teacher, Investigator, and Director of the Bachelor in Elementary Education program from the Faculty of Education. Investigation group member of FORMA. Universidad Mariana, San Juan de Pasto. Correo electrónico:

Fecha de recepción: 28 de enero de 2019
Fecha de evaluación: 20 de marzo de 2019
Fecha de aceptación: 2 de mayo de 2019
Introduction

The purpose of this reflective paper is to analyze and discuss whether standardized testing can measure the quality of education or not based on the ideas proposed by Popham (1999). The first and most important element deals with the homogenization of standardized testing despite the fact that its test takers are heterogeneous and pluriverse.

In terms of standardized testing, Popham (1999) indicated that, “Educators live under a relentless pressure to demonstrate their efficacy. Unfortunately, most communities test teachers’ success based on students’ standardized test results” (p. 2). Therefore, the community is aware that if test scores are high, so is teacher efficacy and efficiency. Yet, if test scores are low, it is the opposite. All of this amounts to a dilemma given that it is impossible to measure every student under the same standards since each one has different conditions and dimensions.

The hermeneutics facing this reality are unfortunate due to the inability to understand how standardized test results affect educators’ teaching process and students’ learning process. Within the educational community, teachers are the most benefited from high test results, but also the least favored by the negative effects of standardized testing. These homogenized exams aim at positioning an educational institute in a particular ranking, which can even indicate the quality of education provided.

Test results are considered a tool that can be used to improve the quality of education because they offer a general diagnosis. Although, the subjectivity and context of the test and where it is applied should be considered. If evaluation is seen as a continuous, permanent, and systematic process, it can be designed, propagated, and implemented with standards that are truly oriented towards quality of education. These tests and standards would consist of commitment and participation from the educational institute and community.

Theoretical Considerations

Before beginning the epistemological debate between evaluation, quality, and education, it is necessary to define and justify each term etymologically. Consequently, these terms and their objectives are constructed considering possible subjectivities in test results and education quality. Furthermore, these terms are defined taking into account the lack of improvement plan ascribed to the institutional context and educational community.

In relation to etymology, we can consider that the term “evaluation” comes from the French word évaluer, and “whose meaning refers to the act of valuing quality or quantity with possible consequences. Quality, from the Latin qualitas, refers to the qualities of something, and education, from the Latin term educare, means to: raise, feed, or instruct” (De la Torre, 2009, p. 20). Once the origin and meaning of these terms are defined, we can affirm
a supposed subjectivity in relation to “quality and evaluation”. Therefore, the relationship between quality and evaluation, with education as its common denominator, is necessary in order to point out the effectiveness of the educational system at all of its levels.

**Education quality, a vague term**

In the field of education, it is difficult to discuss the concept of educational evaluation. It is not an easy area to question and consolidate. In part, this is due to the inability to precisely define educational evaluation in a tangible sense. Thus, it becomes a subjective concept, and at times, abstract as well. In trying to bring meaning to the term, we find support in Egido (2005), who provides conceptual clarity and dynamic theories meant to move on from the problematization of education quality proposed by theorists.

The close relationship between quality and evaluation has been marked by the idea that “if we talk about quality, it is because an evaluation has been conducted implicitly or explicitly. So, if there is an evaluation, it is done with some sort of criteria” (Egido, 2005, p. 2). Evaluation and quality, which also have a close relationship to education, outline the need to work toward education quality with the firm help of evaluation. In doing so, it seeks to obtain the best results in a general sense, and not just by parts or based on certain indicators and results. In this sense, Egido (2006) affirmed that,

Reflections in terms of educational quality cannot be limited to proposing ideas about this concept alone. Rather, it should include reflecting about the procedures used to prove quality, and whether they exist in reality or not. (p. 2)

Up to a certain degree, the concept of “educational quality” could be immersed in a dimension of subjectivity. Therefore, it cannot be reproduced, measured, or precisely quantified. On the contrary, its relativity should be discovered given that there is no one size fits all for knowledge when it comes to education quality, its instruments, or its evaluative procedures.

**Evaluating education quality: origin and evolution**

After referring to the etymological origins of these terms, we take a look at the history of education quality. Its origins date back to the 1930s after the World War II. At that time, the United States sought to improve costs and benefits in production. Quality control founded itself at the beginning of World War II in terms of industrial studies. The need to increase quality became apparent, and statistical quality control became a method to do so in the world of business. The history of evaluation has been around since ancient times, although it picked up strength in the 19th century. During the Pretelyrian period and toward the second and third decade of the 20th century, the idea of evaluation started to pick up in the fields of business and education. Nonetheless, it was Ralph Tyler who coined the term “educational evaluation”, so he can be considered the father of this field. Additionally, evaluation became systematic and professional after Ralph Tyler’s work.

In terms of education, quality and evaluation are necessary in the art of grading systematic, organized processes in order to improve educational systems and the individual’s educative process. Evaluating in order to enhance quality, in all of its aspects and scenarios of life, people, and institutions, can only be obtained through efficiency and effectiveness.

Evaluation should not only be done through standardized tests, nor should quality be restricted to the results obtained from these exams. This is just one aspect of the system and framework of education. Also, national and
international standardized tests should not be the only ones considered. Rather, the entire system should be analyzed so that progress could be made in education quality with the help of evaluation. In order to redo the hermeneutics between quality and evaluation, it is necessary to understand its origins and evolution in education quality. This is the beginning of investigating and consolidating contributions to this area of study.

In actuality, we witness the concern for quality, a crisis which is also being faced by the State. As the administrator, the State must apply public procedures which need to be efficient and effective in the private sector as well. As confirmed by Egido (2006),

From this perspective, the vision of education quality is rooted in purely technical operations which have been inherited from business endeavors of quality. It tends to only consider a set of specifications that should be followed and measured. In reality, as mentioned before, quality is a concept filled with values. Therefore, quality evaluation is not simply a technical process, it is political7. (p. 7)

Quality and education evaluation are not just a technical matter, rather they have a high degree of political commitment in the sense that they can accredit, establish, and execute public policies of the State and not the government. Thus, the concept of education quality, which has been shown to have an etymological basis, will continue to develop its historical evolution. As time passes, education quality will further construct its own epistemological status.

The evaluative mission of standardized tests

The people who design standardized exams are without a doubt very talented. Their objective is to create valid tools that can infer students’ knowledge and skills in a standardized manner. Given this objective, “a standardized test is just like any other test that is administered and follows a predetermined standard procedure of grading. There are two main types of standardized exams: aptitude and achievement tests” (Popham, 1999, p. 3). Parents, educators, and school directors find that standardized tests provide a lot of information in regard to students’ mastery of knowledge and skills.

Often, parents look at their children’s grade and whether it is satisfactory. If necessary, their results may imply additional reinforcement at home. As for the teachers, comparisons are made and improvement plans are established in order to better the learning process. Of course, these plans are meant for the current and future students in order to plan out how to teach content and/or skills related to the subject in a dynamic way.

What does the name have to do with it?

In relation to the definition provided above, Popham (1999) stated that a standardized test has two main types: aptitude and achievements. These types are elaborated, developed, and provided under general principles established for exams, which include validity, trustworthiness, and reliability. Nonetheless, it is necessary to discuss and debate the relationship between trustworthiness and reliability given that there are different interests involved with taking than test than to measure education quality.

From this perspective, Popham (1999) argued that,

Designers of standardized tests to measure achievement have the task of creating an evaluation instrument with many items that can provide valid interpretations of the results. These results
must make reference to standards and the student’s situation in terms of the substantial content that he or she must know

Because of this, test reviewers are always attentive to how the majority of the students answered each of the items provided.

Being able to advance in the labyrinth that is standardized testing and education quality is a challenge. One of these hurdles is showing how these exams do not delve into the reality of the individual, student, or educational institution. Each test taker is different and has multiple particularities in context and hyper-text. Therefore, student conditions, educational establishments, and education quality can never be measured through a standardized test if certain factors are not considered first, such as rural or urban location, official or private entities, and others.

Tedesco (2000) also affirmed that transformative processes in Latin American education adopted a sequence that started with institutional reform. Specifically, this refers to decentralization and the creation of a system that measures results. This coincides with Krawczyk (2002), who mentioned that international organizations are interested in at least three dimensions of sustainability in education reform: politics, finance, and technology. Given the economical proposals established by international organizations, countries must adjust to these three dimensions in order to develop and organize models of educational reform.

Tedesco (2000), Krawczyk (2002) and Cifuentes (2016) all agree that educational reform has been impelled by multilateral organizations. This has led to sponsorship, support, and defense of standardized tests in different countries. In such a way, standardized tests offer the statistical indicators that organizations are looking for as part of their strategies. Eventually, the use of standardized test has become part of governmental policies in different countries. Thus,

the intervention of international organizations, such as the Organization for Economic Cooperation and Development (OECD), the United Nations Educational, Scientific, and Cultural Organization (UNESCO), the joint projects of the Inter-American Dialogue and Research Corporation for the Development (PREAL), the Economic Commission for Latin America and the Caribbean (ECLAC), the Inter-American Development Bank (IDB), the International Monetary Fund (IMF) and the World Bank, among others, are decisive in determining educational reform in Latin American countries (p. 72).

Multilateral organizations and their participation in educational reform can be understood in two ways. The first is through the significant contributions they have done to advance, both dynamically and logically, in a changing and globalized world. This is done within each particular context in order to be at the cutting edge of contemporary society. On the other hand, their intervention has been direct and uncontrolled. Thus, educational reform models are imposed in a constraining manner given that they oppress the State’s models, who also happens to be the one implementing these new reforms. Additionally, there can be penalties and exclusions if these reforms are not implemented and its goals are not accomplished. One example in Colombia comes from the desire to be part of the OECD. Colombia has implemented an infinite amount of reforms to all of its areas, including the education sector, which has meant conforming to the use of standardized tests.

9 Translated by authors.
Measuring the temperature with a spoon

The subtitle may seem illogical, but in effect, it is the thesis that sustains this article. To measure temperature, a thermometer is used. Obviously, a spoon cannot be used to measure temperature because it is a utensil for serving or eating something. This is a paradox where standardized tests cannot measure the quality of education. It is therefore necessary to unravel three realities: education, quality, and national/international standardized tests. Quality is, “specific to the field of education, although quality was an expression used since ancient times. The generalization of its uses had its origin since the 1960s” (Egido, 2006, p. 12). Additionally, Manzano (2015) points out that standardized tests do not measure education quality. This is due in part to the fact that education transcends the barriers of qualification or statistical figures.

Standardized tests should not be used to evaluate education quality. This is based on diverse arguments that prove that it is completely invalid to infer the quality of education based on standardized tests. Education answers to the necessities of its context, and it extends to global and local comprehension. Once again, it is proven that standardized tests cannot measure the quality of education because they are homogenous tests for an otherwise heterogenous population.

What should an educator do?

Popham (1999) indicated that, “educators live under an unmoving pressure to demonstrate their efficacy. Unfortunately, the principal indicator used by most communities to evaluate the success of teachers is through the outcome of the students on their standardized tests” (p. 2). The educative community is up to date on whether the results of standardized tests are high, meaning the performance of the educators is efficient. If test scores are low, this means that teachers are inefficient. This is quite a dilemma because you cannot measure each student with the same stick when each one has different dimensions and conditions.

Teachers display multiple facets in their teaching practices. According to Castañeda (2010), Escudero believed that the teacher must construct his or her own profile and should, “have a profound understanding of disciplinary contents, pedagogical concepts, and the relationship between discipline and subject” (p. 76-77).

Escudero offers a panorama of beliefs, knowledge, capacity, and basic attitudes related to the identity of the teacher. Educators cannot stand against standardized tests, which can be devastating. On the one hand, a teacher cannot dialogue with a standardized test. This would be to lose the epistemological notion of being a teacher by vocation and professional status.

Thus, it becomes necessary to understand each item evaluated on a standardized test and what each one measures. It is indispensable to become familiarized with these tests, and at the same time, defend the idea that quality is much more than quantitative results. What teachers need are instruments that evaluate valuable skills, or significant sets of abilities. In doing so, we can transcend barriers and aid in the construction of citizenship and the transformation of society. Popham (1999) sustained that,

Educators should definitely be held responsible. The education of the future children of a nation is too important to not follow up. Assessing the quality of education through the use of erroneous evaluation instruments is a sub-

---

10 Translated by authors.
11 Translated by authors.
12 Translated by authors.
version of common sense. Although educators need to produce valid evidence about their effectiveness, standardized achievement tests are the wrong tools for that task. (p. 11)

If teachers accept that the results of standardized tests evaluate the quality of education, that is, the teaching process, then the epistemological essence of teaching must be changed. This praxis leads to a horizon different than being a guide, counselor, and motivator. Rather, the teacher becomes part of a learning process that will only be subject to implementing strategies and tools that respond to a standardized test. Therefore, the sense of compassion is lost in the search of aptitudes.

Standardized tests in Colombia: National and International

Standardized tests in Colombia have become a form of evaluation that are normalized and regulated. In fact, the process of standardization dictates the rules of its application and interpretation. Elementary school students from third and fifth grades are already subject to being part of the condition of standardization through the tests known as Saber. These tests were established in the 90’s, and over time, have transcended in their application and results. They have become a determining factor in society.

Table 1. Nationwide Standardized Tests

| ASPECT             | BRIEF DESCRIPTION                                                                 |
|--------------------|-----------------------------------------------------------------------------------|
| SABER              | This test is carried out with students of 3rd, 5th and 9th grade. It began in 1991 with study groups. Between 2002 and 2003, the first sample test was administered with the following baseline subjects: Languages, Mathematics, Natural Sciences, Citizenship and Community Competences, and Social Sciences. |
| SABER 11°          | This test is applied to students in the A and B calendar who are finishing 11th grade. It tests areas such as: Mathematics, Critical Reading, Natural Sciences, English, Social Studies, and Citizenship and Community Competences. The purpose is to evaluate knowledge and abilities of the future university students. |
| SABER PRO          | This test is orientated towards the skills and knowledge of students who are finishing their university studies. It evaluates topics specific to each career. It is administered once a year. Both graduates or citizens can participate in order to establish their understanding in a certain field. |
| Teachers and Teacher Directors | Evaluates the aptitude of candidates who enter the career in education or occupy a position as an educational director. This test is administered annually and seeks to fill positions provided by contests throughout the country. These are specific tests in the field of ethno-education. |

Source: Elaborated by the authors with information from ICFES and MEN.

Students are also subject to evaluation in 7th grade as a partial evaluative element. These results are used to evaluate education quality for 3rd, 5th, 7th, and 9th grade. Likewise, Saber 11 continues to build in strength, becoming a determining factor in granting admission into higher education, especially public education. Due to the diversity of Colombia’s geography, it can be affirmed that these test results do not certify education quality for these grades.

Saber PRO, known previously as ECAES, has become an evaluation which is not taken seriously by students. This is because in order
to graduate, they must present a certificate of attendance. So, their score is not a requirement for graduation, which has led to an inadequate and unconscious preparation for the exam. Since the score is of no consequence to them, students have come to see the exam as unimportant, even though it is required by their academic program and higher education institution.

Colombia is interested in understanding the impact its policies have had on elementary and high school education in relation to other countries. For this purpose, it participates in the Program for International Student Assessment (PISA), the Second Regional Comparative and Explanatory Study (SERCE) and the Trends in International Mathematics and Science Study (TIMSS). These exams are administered on various dates and with different results. The purpose of administering the tests mentioned above is a matter of commitment with multilateral organizations with economic implications, without recognizing the multi-diversity of the contexts in which they are applied.

Table 2. The international standardized tests administered in Colombia

| ASPECT | BRIEF DESCRIPTION |
|--------|-------------------|
| PISA   | This is a program from the Organization for Economic Cooperation and Development (OCDE) and is administered in 58 countries. It evaluates daily knowledge and abilities related to reading, mathematics, and scientific comprehension. They are standardized tests directed towards young adults of about 15 years old who are at least in 7th grade. In Colombia, this trial test was administered in 2005 to 1,720 students from 55 institutions and to 5,250 students from 150 institutions in 2006. |
| SERCE  | This is a project from the Latin-American Laboratory of the Evaluation of Academic Quality (LLECE) from the OREALC/UNESCO in Santiago, Chile. It evaluates basic competences and abilities in the areas of reading, math, and natural sciences which is optional. |
| TIMSS  | TIMSS is an exam from the International Association for the Evaluation of Educational Achievement (IEA). It provides reliable information about the academic achievements in math and natural sciences of students in the United States in 4th and 8th grade. The results are then compared to those of other countries. |

Note. Elaborated by the authors with information from MEN.

From the international standardized tests that are applied in Colombia, the one with the greatest implication is PISA. This is because it can help generate public policies in education and provide a degree of status for the country if the results rise or fall. It is directed towards the government’s need to be a part of the OECD, which incidentally has advantageous and disadvantageous implications in various fields, which include education.

Final Considerations

The field of education quality sets the stage for debate, dissertation, and dialogue of knowledge. In a broad sense, we can analyze its association with its coverage. The commitment of Egido (2005 & 2006), among other researchers, continues in the debate and advance of consolidating the status quo and education quality.

The importance of evaluating academic quality, results, and system effectiveness are a consequence of a process of communication between the participants of the educational process. Through dialogue, systematic coherence is maintained between quality, evaluation, and effectiveness for the achievement of optimal results in education.

Lack of coherence between evaluation and quality can lead directly to the imminent lack of education quality. This is what guides us to question multiple evaluation instruments that could direct progress in educational system. Without these instruments, educational systems
be subject to hopeless improvement, and they would remain fragmented as utopian realities that would never reach their optimal quality.

The intervention of international organizations such as the Organization for Economic Cooperation and Development (OECD), the United Nations Organization for Education, Science and Culture (UNESCO), the joint project of the Inter-American Dialogue and the Research Corporation for Development (PREAL), the Economic Commission for Latin America and the Caribbean (ECLAC), the Inter-American Development Bank (IDB), the International Monetary Fund (IMF) and the World Bank, among others, have permeated different sectors of Colombian society. In Colombia’s case, Saber 3°, 5°, 9°, Saber 11° and Saber Pro have the same dynamic as international tests like PISA, SERCE and TIMSS, whose objective is to measure the quality of education.

The notion that national and international standardized tests measure education quality is subject to discussion. Education quality should transcend standardized tests and their results. They should only be a tool or mechanism used to improve education quality at an institutional level. In other words, it can be used as a projection toward the progress of the teaching-learning process, and not as the final determinant of academic quality.

References

Castañeda, L. et al. (2010). *Enseñanza superior, profesores y TIC: estrategias de evaluación, Investigación e innovación* (1st ed.). Bogotá: Ediciones de la U.

Cifuentes, J. (2016). Políticas Educativas y organismos multilaterales en América Latina *Revista Cultura Científica*, 14, 1657-463X.

Cifuentes, J. (2016). El perfil del docente universitario en la modalidad a distancia. *Revista CITAS*, 1 (2). Retrieved from: http://edu-virtual.ustadistancia.edu.co/Revistas/Revista-CITAS%20Vol-2/

De la Torre, F. (2009). *12 lecciones de pedagogía, educación y didáctica*. Ztapalapa, México: Alfaomega.

Egido, I. (2005). Reflexiones en torno a la evaluación de la calidad educativa. *Revista Tendencias Pedagógicas*, 10, 17-28. Retrieved from: https://repositorio.uam.es/handle/10486/4734

Egido, I. (2006). La evaluación de la calidad de la educación: entre la oportunidad y el riesgo. *Revista Educativa*, 9 (1), 9-23.

Krawczyk, N. (2002). La reforma educativa en América Latina desde la perspectiva de los organismos Multilaterales. *Revista Mexicana de Investigación Educativa*, 7 (16).

Popham, J. (1999) ¿Por qué las pruebas estandarizadas no miden la calidad educativa? Programa de Promoción de la Reforma Educativa en América Latina y el Caribe Grupo de Análisis para el Desarrollo Grupo de Trabajo sobre Estándares y Evaluación. Retrieved from https://www.oei.es/historico/evaluacioneducativa/pruebas_estandarizadas_no_miden_calidad_educativa_popham.pdf

Manzano, M. (2015) ¿Por qué las pruebas estandarizadas no miden la calidad educativa? Retrieved from: https://uls.edu.sv/sitioweb/component/k2/item/326-por-que-las-pruebas-estandarizadas-no-miden-la-calidad-educativa

Ministerio de Educación Nacional de Colombia. (2017). Retrieved from: http://www.mineducacion.gov.co/1621/article-107522.html