

## How do we use educational evaluation to improve?

### Voices from the INEE

Teresa Bracho González

Jorge Antonio Hernández Uralde

Francisco Miranda López

Agustín Caso Raphael

### The Mexican experience

Sonora

Zacatecas

Querétaro

### Analyses and proposals

Guillermo M. Cejudo

Gabriela Pérez Yarahuán

Francisco Abarca Guzmán

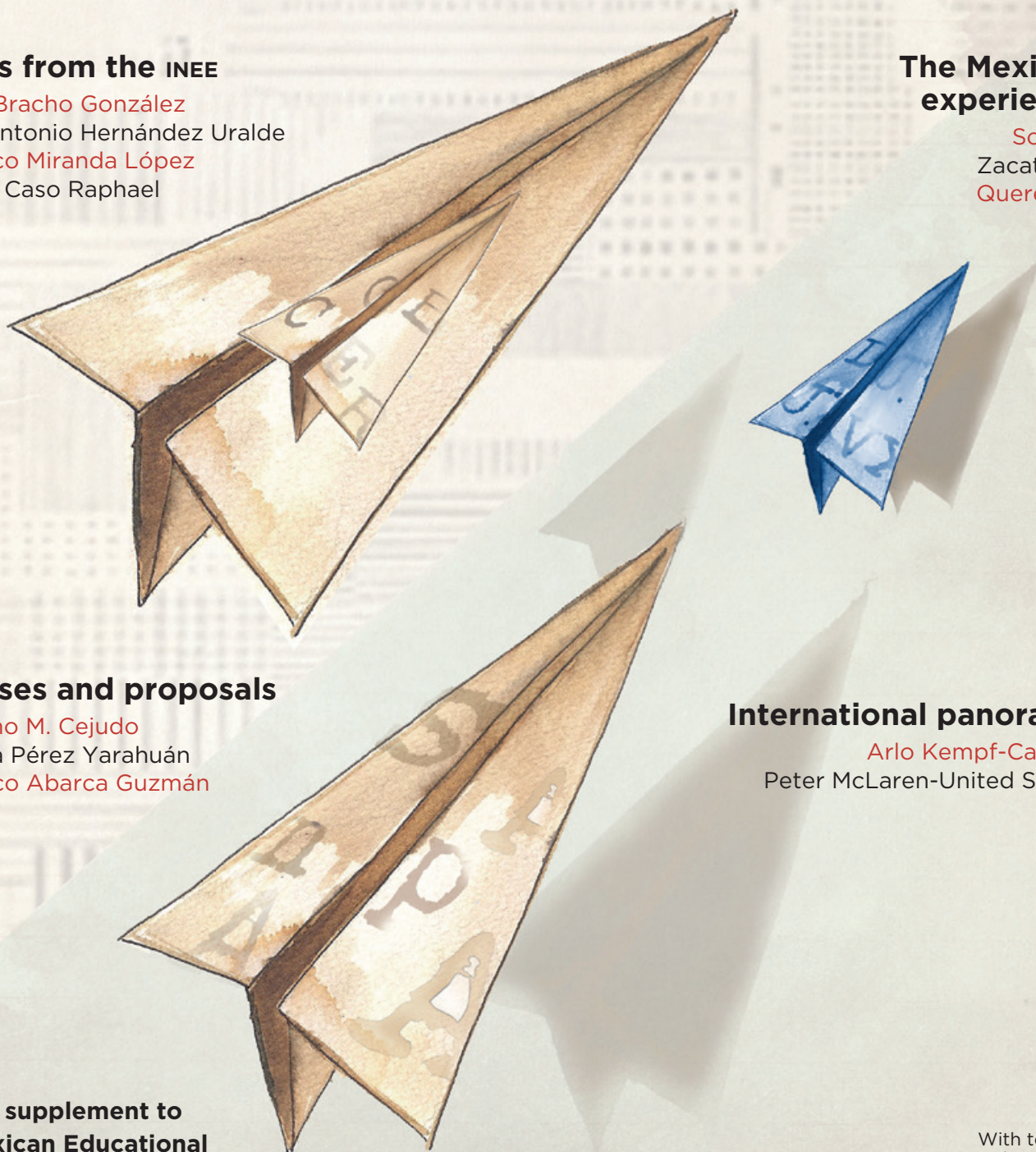
### International panorama

Arlo Kempf-Canada

Peter McLaren-United States

Special supplement to  
the Mexican Educational  
Evaluation Policy

With texts in  
Amuzgo of Guerrero  
and Tzotzil of Chiapas.





Monumento Ecuatoriano

Gestión de Aprendizaje

Organización Escolar

Buena Convivencia

Personal Adecuado

Equipo básico

Materiales de apoyo

Materiales curriculares

Mobiliario

Infraestructura

10

9

6

3

4

5

32

1

# The logbook

National Educational Evaluation Policy Gazette in Mexico

Year 2. No. 4 / March-June 2016

## IN OUR OWN HAND

- 2** **Evaluating to improve: is it possible?**  
Members of the Board of Governors of the INEE

## FROM THE DESK

- 4** **Second year: To build the narrative of evaluation**

## VOICES FROM THE CONFERENCE

- 5** **The Dangers of Educational Extremism: High-Stakes Standardized Evaluation in US K-12 Education**  
Peter McLaren  
Arlo Kempf
- 11** **Education-policy decisions: the link between evidence and values**  
Francisco Miranda López

## FACING OUR CHALLENGES

- 14** **Challenges, aims and work: decisions and educational evaluation in Querétaro**  
José Alfredo Botello Montes
- 17** **An overview of education and evaluation in Sonora: how to overcome the challenges**  
Ernesto De Lucas Hopkins

## OUR VOICE

- 20** **Effective use for improvement; what can we do to make it happen?**  
Teresa Bracho González

## NAUTICAL LETTER

- 23** **A National Educational Evaluation Policy for Promoting Improvement**

## ROADMAP

- 26** **The meaning and use, for the improvement of Education, of the results of the Evaluation of Basic Teaching-Learning Conditions (ECEA) and the Evaluation of Learning Outcomes in Elementary Education (Planea)**  
Jorge Antonio Hernández Uralde  
Andrés Eduardo Sánchez Moguel  
Raquel Ahuja Sánchez
- 31** **The INEE guidelines: building bridges between evaluation, use and educational improvement**  
Arcelia Martínez Bordón
- 34** **2016 Guideline-issuance Program**
- 37** **Information systems on a par with the Educational Reform: innovation for strategic data management**  
Agustín Caso Raphael  
Arturo Cervantes Trejo  
Cecilia Galas Taboada

## IN THE CLASSROOM

- 40** **Can the project for certifying educational quality via ISO 9001-2000 improve education in Zacatecas?**  
Adolfo Castruita Monreal

## DOSSIER: USE OF INFORMATION, TWO PROPOSALS

## OTHER PERSPECTIVES

- 42** **Gazette No. 4. How to foster the use of evaluations?**  
Guillermo M. Cejudo  
Francisco Abarca Guzmán
- 46** **The Use Radar**  
Gabriela Pérez Yarahuan

Visit the *Gazette* digital version in the INEE web at: [www.inee.edu.mx](http://www.inee.edu.mx).

*National Educational Evaluation Policy Gazette in Mexico*, Year 2, No. 4, march-june 2016, is a publication of the National Institute for Educational Evaluation by the Educational Policy and Regulations Unit, under the care of the General Direction of Guidelines for the Educational Improvement. Barranca del Muerto No. 341, Col. San José Insurgentes, Del. Benito Juárez, C.P. 03900, México, D. F. Tel. (55) 62.70.66.00.

Coments: Laura Athié, Responsible editor ([gacetapnee@inee.edu.mx](mailto:gacetapnee@inee.edu.mx) / [lathie@inee.edu.mx](mailto:lathie@inee.edu.mx)).



**INEE**  
<http://www.inee.edu.mx>



**Twitter**  
[@INEEMX](https://twitter.com/INEEMX)



**INEE Facebook**  
<http://goo.gl/axitPa>



**INEE Youtube**  
<http://goo.gl/fHRDvC>

## Evaluating to improve: is it possible?

MEMBERS OF THE BOARD  
OF GOVERNORS OF THE NATIONAL  
INSTITUTE OF EDUCATIONAL EVALUATION  
(SPANISH ACRONYM: INEE)

*In order for the link between evaluation results and educational improvement to occur, it is necessary that both education-authority decision-makers at all levels and also teachers develop the competencies that will enable them to understand the contents and scope of the said results [...], that their value and importance be acknowledged, that they be considered kind of accountability, and that the technical expertise of the evaluating bodies and the professional role that they play be recognized.*

Governing Document  
the National Policy for  
the Evaluation of Education

Education is a human right and plays a decisive role in developing people's abilities, skills and capacity to enjoy life,<sup>1</sup> and, for its part, educational evaluation is a tool designed to improve education and thus to make good on the right to universal high-quality education.

But despite all the good intentions and high expectations, the truth is that there is no automatic link between evaluation and improvement. Such a connection may fail to occur when evaluation is not pertinent or relevant, or not properly focused, yielding results that are not germane, or when it is not timely or it takes so long to give feedback on its results that the crucial opportunity for them to have an impact is lost. Also, if there is not adequate communication about an evaluation, resistance to it might arise.

In all cases, the main thing is that people must know about, and use, evaluation if it is

to drive improvement; if its results just sit on the shelf, their likelihood of informing decision-making will decrease, and hence they must be disseminated as necessary, via the right means and to the right and people, and suitably used for the purposes that their nature and technical features equip them to be used, based on awareness of their consequences and certainty that they will serve to foster the improvements in education that the National Institute for the Evaluation (Spanish acronym: INEE) wants to achieve in the context of the Educational Reform in Mexico.

We at the INEE are convinced that evaluation findings constitute essential input for the creation of the National Educational Policy, and that for the latter to be optimally formulated, the voices of all those involved in education, both nationwide and at the local level, must be heard if we are to get a clear idea both of the realities and problems of the education system, and also of the solutions needed by a diverse, multicultural country such as ours. Since evaluation is a mere means, rather than an end in itself, we must not only create good evaluations, but also use them, and therefore the INEE's main aim is to ensure that evaluation leads to improvement – i.e. that it informs decision-making aimed at making education more equitable and raising its quality.

With the above aim in mind, in the fourth edition of the *Gazette of the National Evaluation Policy in Mexico*, we seek to ascertain: (i) what the uses of the evaluation that is being produced by the National Evaluation System (Spanish acronym: SNEE) – primarily by the INEE, but also by the education authorities – are; (ii) what the biggest achievements of – and challenges facing – the said evaluation are; (iii) whether the said evaluation is indeed used, and, if it is, by whom and how; (iv) what we can do to foster more use; and (v) if the said evaluation serves to improve education.

More precisely, this edition of the *Gazette* sets out to find out (i) what types of evaluation findings tend to be used for purposes of policy and program design; (ii) what the biggest challenges are regarding the dissemination and use of the said findings; (iii) what needs to be done to ensure

that decision-makers and others involved in education access the said findings and use them to make improvements; (iv) how the results of the tests pertaining to the National Plan for the Evaluation of Learning (Spanish acronym: Planea) and the Evaluation of Basic Teaching and Learning Conditions (Spanish acronym: ECEA) can be used to plan improvement paths in schools and school districts; (v) what needs to be done to foster greater use of the evaluations carried out by the INEE for planning purposes; (vi) whether the design of the evaluations needs to be changed; (vii) whether the evaluation findings need to be disseminated in a different way; (viii) who we need to liaise with to foster greater use of the said findings; (ix) what features an information system needs to possess to ensure that it is optimally used by decision-makers; and (x) what proposals the model for the creation of guidelines needs to include in order to improve education.

Based on the aforesaid questions, in this edition of the *Gazette* we include some local viewpoints from the states of Querétaro, Sonora and Zacatecas, along with two proposals by Mexican specialists, as to ways in which evaluation findings can be used. We also include an article by Peter McLaren and Arlo Kempf, two of the most influential theorists in the field of critical pedagogy, who talk about the limits of standardized evaluation in the light of the experience with such evaluation in Canada and the USA.

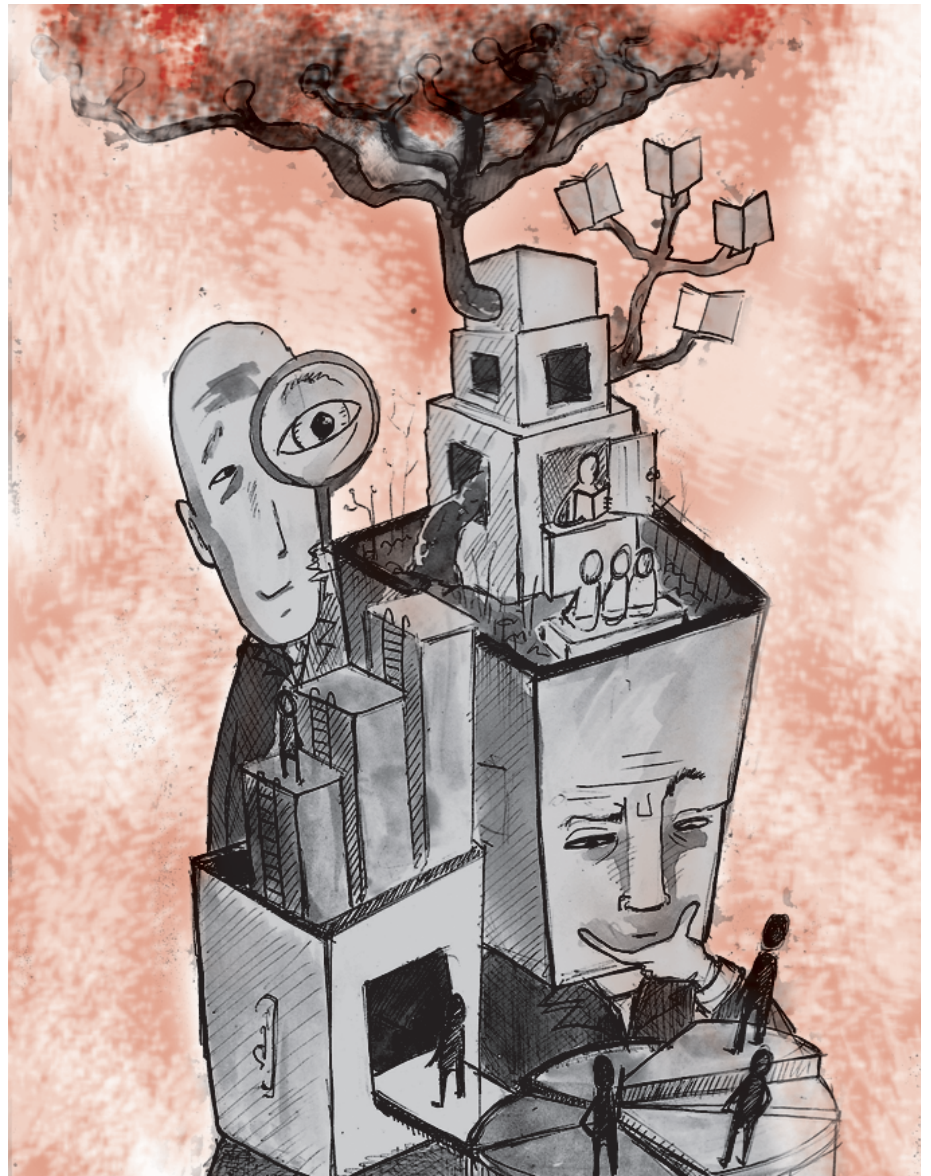
The *Gazette* also includes contributions by the heads of various INEE units, by Teresa Bracho, a member of the INEE's Board of Governors, and by other INEE officials who, with regard to the necessary connection between evidence and fundamental values, describe the progress achieved in our own evaluation system, make proposals for the use and dissemination of evaluation findings, and present an analysis of the Planea and ECEA results. These reflections are accompanied by a text that talks about the creation of guidelines that can foster the use of evaluation to improve education via the influencing of education policy.

This year's first edition of the *Gazette* also includes a section devoted to the National Educational Evaluation Policy (Span-

ish acronym: PNEE) that will continue to appear in all the subsequent editions, since the latter is the benchmark for the development of evaluation aimed at educational improvement that the INEE wants to promote. Given the importance of this topic, we are also including a special supplement devoted to it, developed by all the members of the INEE's Board of Governors, two experts from UNESCO's International Institute for Educational Planning in Buenos Aires, and the David Rockefeller Center for Latin American Studies, who talk about the achievements of the PNEE and the challenges it faces, based on an analysis of its recently published Governing Document.

It should not be forgotten that 2016 is a crucial year for education in Mexico. We are exactly halfway through a 6-year presidential term of office marked by an educational reform that gave rise to ambitious evaluation projects aimed at professionalizing teachers and also set up the National Educational Evaluation System (Spanish acronym: SNEE), which, possibly the only system of its kind in the world, is outstanding for vigorously promoting the development of measurement and evaluations, not only of learning outcomes, but also of various components, processes and results, as part of a systematic effort to link evaluation to different regulatory and technical endeavors in order both to ensure that it is transparent, equitable, fair and objective and also to foster cooperation at the institutional level among the different education authorities and promote strategies for influencing education policies and programs, improving education services, raising the latter's quality, and achieving more equitable results.

In 2015, important changes were achieved in the evaluation of student achievement and learning outcomes via Planea, which was implemented for the first time in the final years of primary, lower-secondary and upper-secondary education. We carried out a *Consulta Previa, Libre e Informada a Pueblos Indígenas sobre Evaluación Educativa* ("Preliminary, Free, Informed Consultation of the Indigenous Peoples") about the education received by the indigenous peoples and the evaluations carried out in their schools, as well as ascertaining the conditions in Mexi-



co's primary schools via the findings of ECEA. At the end of the same year, based on the results of the competitive examinations for entry to the teaching profession, the INEE issued its first guidelines aimed at improving initial elementary-level teacher training.

Thus it bears asking whether all the work described above will have an impact on the formulation of educational policy, when this will happen, and what needs to be done to ensure it.

Our guest contributors to this edition of the *Gazette* reflect not only about educational evaluation per se, but also about its achievements, the challenges that must

be overcome to ensure its use, and its potential for informing educational policy and achieving improvement.

We would like the debate initiated in this 4<sup>th</sup> edition of the *Gazette* to be ongoing and thought-provoking while also opening the door to input resulting in new ideas and practices that consolidate the new wave of improvement-focused evaluation that is underway in our country.

1 See: Amartya Sen. (1999). *Development as Freedom*. [www.uio.no/studier/emner/matnat/ifi/](http://www.uio.no/studier/emner/matnat/ifi/)

## Second year: To build the narrative of evaluation

The art of recounting events, experiences, tragedies, the enormous problems we face and how we deal with reality is as old as the world and we practice it via the printed page, emails, blogs, video clips, books, reports with graphs and statistics and documentaries, telling stories that give rise to comments, bring together shared attitudes, answer questions and clear up doubts, doing so by means of words that combine to create a narrative, captivate our audience, make it remember and, at the end of the day, leave it with a new experience, some queries, or something to reflect on.

Today this way of communicating things has been -and is still- known as storytelling, which, as explained by Mo Yan, the winner of the 2012 Nobel Prize for Literature, whose name in Chinese means “don’t talk”, simply means letting the rest of the world know about things:

I know that nebulous terrain exists in the hearts and minds of every person, terrain that cannot be adequately characterized in simple terms of right and wrong or good and bad, and this vast territory is where a writer gives free rein to his talent. So long as the work correctly and vividly describes this nebulous, massively contradictory terrain, it will inevitably transcend politics and be endowed with literary excellence.<sup>1</sup>

The same is true of the *Gazette*, which enters its second year of publication today, resolved not to be silenced, but rather to go on telling the story, to find that grey area in order, on each of its pages, to tell the story of evaluation, to talk about the educational reform that began in 2013, and to publish the articles of its contributors who, also part of the reform, have expressed their opinions about it in the course of five editions.

This way of explaining complex ideas in simple words is, precisely, a form of storytelling – the transmission of a message that

opens people’s eyes and gives us a chance to tell out stories after listening to those of other people.

Beyond numbers there are words, and, beyond words, people – the teacher who tells us that he does not feel he has been represented in the *Gazette*, the education minister who talks about his plan, the specialist who, in these pages, proposes an evaluation model aimed at solving the big problems that the National Educational Evaluation System has identified in Mexico’s education system and seeks to solve. Each author-protagonist reads and recounts, thus contributing to a shared account of the endeavor to implement a countrywide policy...and, in their emails and responses to our surveys, the readers of the said account have expressed the following needs:

- 1. We’d like to have our concerns addressed.** In response to this request, as of this year, *the National-Education-Policy Gazette will be including, in its “Our Readers” section, a dialogue with those who seek clarification.*
- 2. We need to get a better understanding of the National Educational Evaluation Policy.** In answer to this comment, in this 4<sup>th</sup> edition of the *Gazette* and in subsequent ones, we will be including a policy report, with graphs, about the decisions made -and the projects carried out- in relation to the seven main components of the National Educational Evaluation Policy.
- 3. We’d like to share our opinions and suggestions.** In response to this concern, the *Gazette* is becoming interactive and will be including two capsules featuring video interviews with specialists on its microsite in each edition, as well as spac-

es for readers’ comments and forums for exchanging ideas and opinions: <http://www.inee.edu.mx>.

- 4. We’re looking for information as a basis for making decisions and would like to know what responses are received to the questions posed in each edition.** Since people need input in order to make decisions, twice a year we will be including a special supplement like the one to this 5<sup>th</sup> edition, which is entitled *The National Educational Evaluation System and the National Educational Evaluation Policy: Progress and Viewpoints.*
- 5. We decision-makers need to have a say; there are a lot of us, in different spheres of action.** In response to this comment, the *Gazette* is publishing the following invitation and will be welcoming your input so as to ensure that all those people who work in schools, research centers, institutes and supervision and evaluation offices are also represented.

Specifications for contributions: In digital form (Microsoft Word file), double spacing between lines. Arial-bold 12-point font. Maximum size: 12 single-sided A5 sheets, to be sent before the closing date to: [lathie@inee.edu.mx](mailto:lathie@inee.edu.mx). All contributions will be reviewed by the Editorial Board.

By working together to write this account of evaluation, which reaches 22 countries and thousands of readers, we are, in the words of Mo Yan, “reading about society and life, which, together, constitute a big, invisible book” where, in just a few pages, people can understand the whys and wherefores. “What I did,” says Yan, “was to tell my stories in my own way”. €

<sup>1</sup> Mo Yan, *Sorytellers*. Nobel Speech. December 7<sup>th</sup>, 2012. © NOBEL FOUNDATION 2012. <https://goo.gl/doGiZC>

**If you want to contribute to the *Gazette*, please take note of the following information:**

Edition	Topics	Closing date
<i>Gazette</i> No. 5 (July-October, 2016)	Context, diversity and educational evaluation: towards fair, inclusive education?	May 15 <sup>th</sup> , 2016
<i>Gazette</i> No. 6 (November 2016-February 2017)	Evaluation and local capacities: can federalism be rebuilt?	August 15 <sup>th</sup> , 2016

VOICES FROM THE CONFERENCE  
THE SPECIAL GUEST

## The Dangers of Educational Extremism: High-Stakes Standardized Evaluation in US K-12 Education

“We write this with humility, as a modest missive to education stakeholders in diverse Latin American contexts [...]. These are challenging times for policy makers, and the question of how we use evaluation and assessment in our schools is fundamental to our vision and philosophy of education,” say Peter McLaren, the father of the Critical Pedagogy, and Arlo Kempf, teaching and learning expert, in this analysis about what they call the Global Educational Reform Movement.

**PETER MCLAREN**  
College of Educational Studies, Chapman University  
University Of California at Los Angeles (UCLA)  
[peter.mclaren1@gmail.com](mailto:peter.mclaren1@gmail.com)

**ARLO KEMPF**  
Coordinator of Secondary Education, Master of Teaching Program  
Department of Curriculum, Teaching and Learning, OISE/University of Toronto  
[arlo.kempff@utoronto.ca](mailto:arlo.kempff@utoronto.ca)

### Introduction

The past fifteen years have seen a radical overhaul of education in the United States. Among the most significant elements of this transformation has been the expansion and acceleration of the use of high-stakes standardized testing. Despite major concerns about accuracy, student well-being, and social and economic inclusion from various stakeholders the use of testing is currently at an all-time high, and correlates to a host of negative consequences for K-12 students



and other education stakeholders in the US. While recent education reforms, including the Every Student Succeeds Act (ESSA), may be a move away from one-size-fits-all approaches to understanding student success, there remains a complex web of local, state, and federal tests and testing mandates. Further, the data driven culture of so-called accountability persists throughout US education, with students, teachers, administrators, schools, and districts increasingly understood through the narrow prism of testing results.

Recognizing the diversity of educational realities across Latin America at the national, subnational, and intranational levels, this article offers a humble and general caution for Latin American policy makers and stakeholders, with reference to the hazards of the overuse of standardized assessment. The first section below introduces the current context of high-stakes testing in the US, with a look at the quantity, frequency, and duration of testing. The next section situates US K-12 standardized testing practices in international perspective. The third section looks at the politics of testing as well as the surrounding myths and contested rationales underpinning high-stakes testing such as data driven accountability. The article then turns to consider implications for education in Latin America. A short conclusion follows.

### Current Context of US High-stakes Standardized Testing in US K-12 Classrooms

In the US, innumerable and wide-ranging standardized tests (ST) are used to measure the learning, understanding, and aptitude of children. Some are diagnostic, indicating they are geared to assess students' knowledge and ability in order to inform subsequent teaching activities and foci. These may not accurately be called high-stakes, as they generally do not impact whether students advance or not, and student performance on these assessments is unlikely to impact school funding, school reputation, or teacher salary. We can thus set these aside in this discussion, and focus exclusively on ST that is not diagnostic, and on those tests which are used across US districts and/or states.

In the US, the quantity of testing is different from state to state and from district to district. Many states administer standardized tests in the first and second grades, however, many begin ST as early as kindergarten

(roughly ages 3-5) and continue all the way through to high school graduation (roughly ages 16-18). In addition to state- and district-mandated testing, the Common Core State Standards Initiative (CCSSI) sets out guidelines for curriculum, delivery, and assessment (including ST) in states that have adopted the Common Core (a strong majority as of January 2016). In December of 2015, US President Barack Obama signed the Every Student Succeeds Act (ESSA) which is an overhaul of the previous No Child Left Behind Act (NCLB). While it is possible that CCSSI and ESSA may ultimately reduce the quantity of standardized testing, it is clear that current use of these high stakes assessment is unprecedented in US history. A study by the American Center for Progress (2014) found that students were writing up to 20 standardized tests per year, with an average of ten per year in grades three through 12.<sup>1</sup> These figures support research conducted by the American Federation of Teachers.<sup>2</sup> Additionally, Kempf's (2016) study of teachers in Chicago found teachers administer an average of 11 standardized tests per year, while *Education Week* suggests that in many urban schooling districts, students are writing an average of 113 standardized tests between pre-k and twelfth grade.<sup>3</sup> Although there are not many US jurisdictions in which students write fewer than five standardized tests per year, the research mentioned above indicates there are many in which students write up to 20 or more on a yearly basis.

Generally speaking, quantitative increases in the use of standardized tests correspond to quantitative increases in classroom time allocated for test preparation. A 2014 *New York Times* article described the situation in Florida where "many schools this year will dedicate on average 60 to 80 days out of the 180-day school year to standardized testing (ST). In a few districts, tests were scheduled to be given every day to at least some students."<sup>4</sup> Former US Assistant Secretary of Education Diane Ravitch suggested that in late 2013, many districts were allocating "20 percent of the school year" to test preparation.<sup>5</sup> The American Federation of Teachers study found that each year, approximately 20 days of instructional time are used in test taking, administrative activities and test preparation.<sup>5</sup> It is safe to say that on the low end in the United States, ten percent of class time is used for activities related to ST, with a number of jurisdictions devoting 20 percent or more.

### US K-12 Evaluation in International Comparative Perspective

The United States is an international outlier not only as far as the quantity of standardized testing used, but also as far as the degree to which standardized tests are used to determine educational success (on numerous levels). Students in the US write more standardized tests more often than most students around the world. These standardized evaluations are used not just to understand student success and ability but also to evaluate teachers, schools, districts, and states under the banner of data-driven accountability. With the United States at one end of the spectrum as far as testing, Finland is at the other. The Scandinavian nation uses only one standardized test, at the end of secondary school. Sixteen-year-old Britons, by contrast, write a series of high-stakes tests (approximately 15–20), which determines their path forward in secondary school and thereafter in potential opportunities to continue to college or university.

The Program for International Student Assessment (PISA) administered by the Organization for Economic Co-operation and Development (OECD) offers educational measures of 65 countries around the world, and provides a broad international perspective on testing usage. One measure relevant for understanding testing in comparative perspective is PISA's look at the use of examinations at the lower and upper secondary levels. According to the OECD, lower secondary generally refers to schooling for students aged 10 to 16. It begins between "the ages of 10 and 13 and ends between the ages of 13 and 16. It is compulsory across OECD and partner countries. It marks the end of compulsory education in a number of countries."<sup>6</sup> Upper secondary is typically two to three years, with students generally entering at 15 or 16 years of age.

While PISA does not offer data on ST at the elementary level (what the OECD calls the pre-primary and primary levels) findings on the use of ST for older students are important. PISA differentiates between assessments and examinations. Assessments "take stock of students' performance in order to make decisions on future instruction or to summarise performance for information purposes" and "do not have direct tangible consequences for students."<sup>7</sup> Alternately, examinations "determine students' pathway to upper levels of education (e.g. the transition

from lower to upper secondary school), selection into different programmes (e.g. into vocational or academic programmes), or selection into university programmes."<sup>8</sup> While all of the 65 countries have at least some assessment and/or evaluation at the lower and/or upper secondary levels, there is significant variation, and the United States is among a very small number of countries that uses national/state and additional examinations at both the lower and upper secondary levels. Although the US uses relatively few high-stakes entrance exams for qualification for secondary and postsecondary opportunities, the intensity of testing beginning as early pre-K is extraordinary and internationally unparalleled.

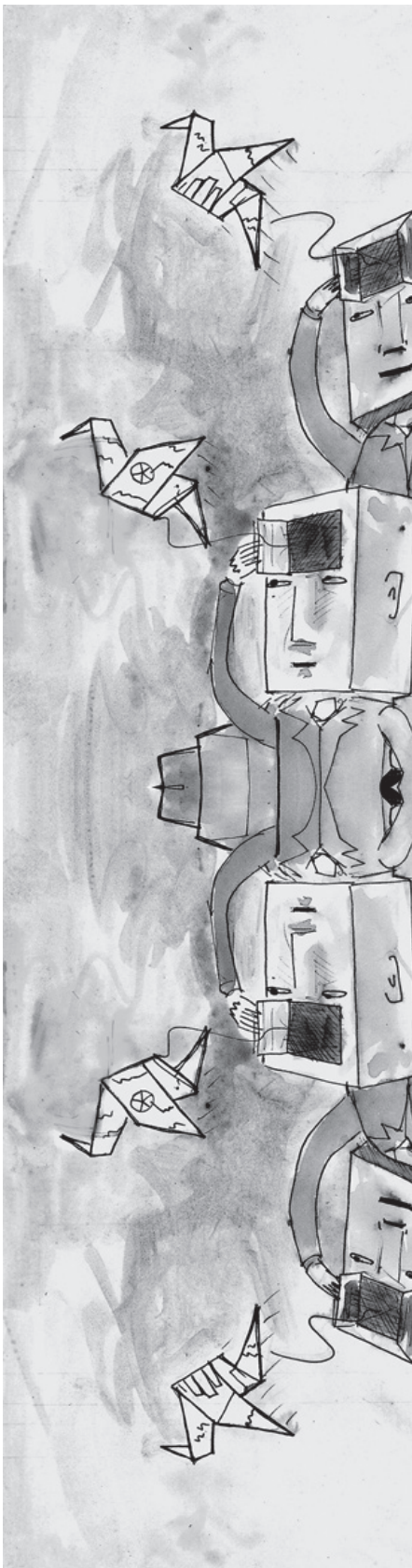
### The Myths of Accountability, Impartiality, and Equity: The Challenges and Rationale for Standardized Testing in the US

Among the most common justifications for the use of standardized testing in the US are that ST offers an impartial measure of students' accomplishments, skills, and understandings in a given subject area or areas; that these impartial methods provide an equitable educational opportunity for low income and racial minority students (moderating the educational impacts of race and income); and that such measures provide accountability to taxpayers as far as a return on taxpayer dollars. It is worth discussing each of these pieces separately here before moving on.

#### Impartiality

The most noticeable barrier to impartiality, or objectivity, is inaccuracy. In the US, countless incidents of cheating have raised serious questions about the validity of the use of standardized testing results. Cheating, however, is not the biggest issue as far as objectivity and impartiality and ST. In order for standardized tests to be objective they have to be clear (what they claim to measure must be what they are designed to measure, and as well correspond exactly to what they do measure); they must be accurate (based exclusively on classroom and school learning, with no interference); and they must be reasonably value-free, safeguarding that culture and income do not privilege or handicap any students in preparing for or succeeding on the test (nor any teacher in preparing students for it). To the question of clarity (whether





standardized tests measure what they claim to measure, well-known US critic of ST Alfie Kohn argues that “[n]orm-referenced tests were never intended to measure the quality of learning or teaching” and suggests “the main objective of these tests is to rank, not to rate; to spread out the scores, not to gauge the quality of a given student or school.”<sup>9</sup> Norm-referenced testing in schools compares one student’s performance to that of all others taking the same test. This is different from evaluating what a student has learned or what a student knows, and additionally tells us nothing about changes in an individual student’s academic progress.<sup>10</sup>

Numerous academic researchers have argued that ST frequently fails to meet sufficient standards of both reliability and validity.<sup>11</sup> Put simply, we cannot ensure that ST data are reliable as far as understanding the academic ability of individual children. CRE-ATE, a group of researchers from Chicago, explains the difference between reliability and validity using the metaphor of two broken scales. One gives you a different weight each time you step on it (unreliable) and the other is always ten pounds over (it is reliable but invalid).<sup>12</sup> American academic William Popham suggests that validity is the most important component of any evaluation.<sup>13</sup> This is a criterion, however, that standardized testing has not met in the opinion of many scholars (of different political persuasions) for the last 25 years.

### The Myth of a Level Playing Field

Kempf (2016) notes that a great deal of standardized evaluation unavoidably assesses what students have learned out of school. For example, the third grade student who spends her Sundays in a private chess camp will come to school with an advantage over a classmate who spent his Sunday in Bible school. Each will have been gaining knowledge but one is learning ideas, language, and analytical processes that are validated and recognized as legitimate in-school knowledge. As far as testing, the issue here is not which weekend learning is better but instead that one student is coming to school with a tool for success, a tool to which most students have no access. Our use of standardized tests may then equate this advantage with intelligence.<sup>14</sup>

Standardized testing thus not only fails to account for economic, cultural, and other differences within student populations, it

can also transform social difference into academic advantage and disadvantage. When test results are used to judge schools, districts, and increasingly teachers for things over which they may have limited or no control, the dangerous role of testing becomes clear. Students, teachers, schools, and districts are understood in either a positive or negative light based, sometimes exclusively, on ST scores. Successful results often correlate to higher incomes among parents, while by related contrast, low test scores correlate to higher rates of poverty.<sup>15</sup>

The Trends in International Mathematics and Science Study (TIMSS), released in 2011, reported wide differences between the reading and math performance of eighth grade white, African American, and Latino students.<sup>16</sup> Additionally, there had been no change in the gap between the performance of 17-year-old African American and white students on the National Assessment of Educational Progress in both reading and math since 2004.<sup>17</sup> Using graduation rates as a measure of school success, there was a 20 percent decline among students in high-poverty schools, while rates for students in low-poverty schools remained the same.<sup>18</sup>

### Accountability

As Kempf has argued elsewhere<sup>19</sup>, the driving rhetorical force in the push for ST is the idea of accountability. It is a compelling discourse. Who would disagree that teachers should be held to high standards, that students should be well educated and cared for, that money from taxpayers should be carefully spent, that public servants with inadequate performance should be forced to improve or lose their job, or that we should have clear tools for understanding how well our education systems are working? These common sense arguments make up simple and seductive justifications for standardization, however, there is no proof ST (or standardization of education more generally) accomplishes any of these objectives. Excessive use of standardized testing can professionally disadvantage teachers and administrators who work with student populations with low test results. Gauging student and teacher ability through ST can lead some teachers to prefer teaching only those student populations who are likely to be high performing according to test results, and indeed push them away from serving students who are not.

After over fifteen years of increased use of high-stakes standardized testing, we are no closer ensuring accountability to the public. Further, as Kempf has argued<sup>20</sup> ST has failed to produce gains such as better support of students with diverse learning needs, more inclusive curriculum, and increased teacher access and application of research into teaching and learning. Indeed, standardization has largely hindered rather than promoted these improvements. There is a painful gap, then, between blanket calls for accountability and educational improvement. Whereas there is no doubt that education in the United States faces many problems, there is no proof that standardization will fix any of them.

From a wider international perspective, Finnish educationalist Pasi Sahlberg suggests we are now witnessing a “Global Educational Reform Movement” (GERM) and argues systems that have experienced increased standardization (specifically a move toward test-based accountability, standardization, school choice, and human capital model thinking) have seen a corollary decline in academic performance.<sup>21</sup> Citing 12 years of OECD/PISA data from 2000 to 2012, Sahlberg traces the decline in mathematics learning outcomes in the United States, England, Canada, Australia, New Zealand, the Netherlands, and Sweden—all jurisdictions, which have embraced the GERM approach, to albeit different degrees.<sup>22</sup> In the United States in particular, over 25 percent of students fail to meet OECD/PISA’s baseline for math proficiency, and nationally the United States ranks 27<sup>th</sup> among OECD countries.<sup>23</sup> While standardization does not correlate with academic improvement, OECD/PISA data on social and economic equality tells us that improvements in social and economic equality correlate with improved performance on OECD/PISA evaluations,<sup>24</sup> a conclusion supported as well by the work of Wilkinson and Pickett.<sup>25</sup> The United States, which has the same levels of inequality today as it did in the 1790s,<sup>26</sup> also lags in its educational performance.

Internationally renowned educationalist Andy Hargreaves muses that: “accountability is what remains when responsibility has been subtracted.”<sup>27</sup> As Kempf has suggested<sup>28</sup>, the push for accountability deprofessionalizes teachers’ work; establishes competition at the center of our education system; and alienates students and parents

from the constructive, experiential, and social elements of learning. Internationally, the United States is a radical standardized testing outlier, testing more and more often than anywhere else. Although the US uses fewer make-or-break examinations (such as middle school or high school entrance exams) than some other places, the intensity of standardized testing (beginning as early as Pre-K) sets it apart from many other education cultures. Most teachers assess students very frequently, and the vast majority of that assessment would make no sense expressed as simply as a number. Often lost in the standardized classroom is the context, the fecundity, and the unique challenges, strengths, and opportunities which that are taken into consideration in the course of rich and useful assessment and evaluations. That education is a political activity coeval with economic life is by now an accepted commonplace in the nomenclature of educational theorists. Given the grim reality of persistent poverty in the United States—and the hard truth that a large segment of the new jobs being delivered by the transnational capitalist economy pays little more than poverty-level wages—even if poor parents can raise their children like middle-class parents, their chances of escaping poverty through education is far from assured. Seventy-four percent of the entering class at top-tier colleges and universities come from the ranks of the economically well-off; that is, from the top quartile of families by income group.<sup>29</sup> Most of the economic growth in the U.S. is benefiting the rich, and low-wage jobs are becoming a permanent part of the U.S. economy. Already, one in four Americans has a job that pays poverty-level wages. Indeed, fighting for the right to have a decent and remunerative job would do more for educational reform than current policies involving testing.<sup>30</sup>

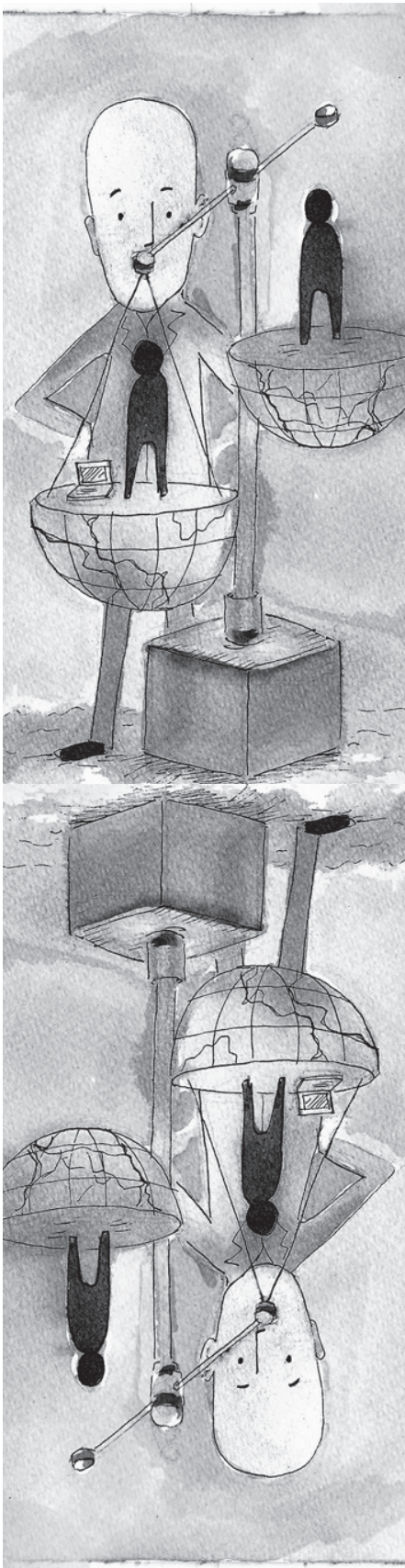
What are the moral and educational implications surrounding the use of multiple-choice tests that rely on memorization and test-taking skills in our contemporary world; a world where electronic and digital media shape our neural networks, mediate our perceptual emotional networks, and reorganize our senses and neuropsychological structures as well as speed up our “orienting response” regardless of content—all which combine to create attentional and memory difficulties later on in school?<sup>31</sup> Compound this dilemma with the fact that many stu-

dents from poor families already experience unhealthy levels of stress hormones, which can impair their neural development at a very early age.<sup>32</sup> What are the implications of high-stakes, standardized tests in a digitalized world where school administrations attempt to increase the RAM and gigabytes in classroom computers rather than find ways to increase the shrinking attention spans of students that result from intense computer use as well as other factors? Economic inequality can affect early childhood development in many areas including brain development, educational achievement, and mental health. The holistic burden should thus not be placed solely on the shoulders of teachers, since it is impossible for even the best teachers to overcome these obstacles facing their students despite the excellence of their instructional programs and the high quality of their pedagogical skills. The struggle for educational reform *is* the struggle for economic justice. There is no other way to put it.

### Considerations for Latin America

The greatest impediment to educational success and prosperity is not a lack of good teaching or inadequate testing and accountability but rather economic and social inequality. This is global truth with particular resonance in Latin America. If this is the case, then following in the footsteps of the United States as far as high-stakes testing will likely worsen national and subnational issues rather than help to solve problems facing education stakeholders in Latin America. The solution to educational reform and the quality of social life in general in the US relates powerfully to economic rights—indeed to rights that are not always attached to educational reform. This is equally true of countries throughout Latin America. Education reform in democratic contexts must ultimately enhance educational opportunities for the popular majorities—it must recast society through the promise of political democracy by extending educational democracy into the realm of the social and economic through a political process of popular self-emancipation. While, short of this lofty goal, there may be many efforts that can be made to improve education, administering high-stakes testing is not among them.

Testing is clearly linked to the global marketplace, educating for specific forms of labor power that will serve the interests of the transnational capitalist economy. However, critics such as John Marsh (2011) make



the case that education should be treated as a political—not a market—phenomenon and we believe that there is merit to this argument. Marsh notes that the US does not generate many more poor people than other countries. European countries achieve lower poverty rates because they provide more social programs aimed at the poor and unemployed. Without government programs, Sweden would have 26.7 percent of its population living in poverty. With a strong social safety net, however, including policy concentrations on education, the poverty rate is 5.3 percent.<sup>33</sup> To be sure, education helps some people enter the labor market, and indirectly might create a few more jobs, but what we need is direct job creation, higher wages, and better redistribution programs.

Marsh notes, citing various researchers, that among children whose parents have identical levels of education, those children who lived in unequal countries performed worse on tests of adult literacy. Children of parents with college degrees in general perform the same, whether they live in Finland (one of the most economically equitable countries) or the US (one of the most economically inequitable countries within the so-called developed world). But children in the US whose parents only attained high school will perform worse on literacy tests than children in Finland whose parents have the same education as their US counterparts. This is because economic inequality affects the quality of family life, in areas of health, security, rates of substance abuse, etc.<sup>34</sup>

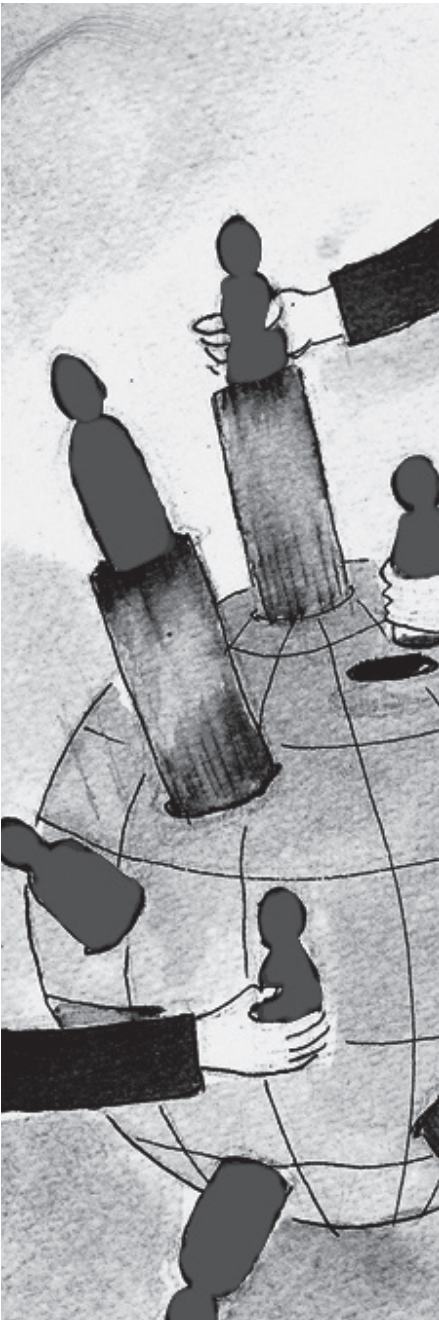
Education has been made the only available means of addressing injustices that arise from economic disparity, and to us this constitutes one of the worst crimes of capitalism. In our view, the answer to improving education lies in creating more economic equality and equity. This is a lesson that all capitalist contexts (in Latin America and elsewhere) need to learn, as the United States seems oblivious to it. Focusing on testing and accountability will only further exacerbate already dire conditions for the marginalized and working-class communities throughout Latin America.

### Conclusion

In this article we have attempted, albeit briefly, to highlight and survey the perils of the current approach to education and student evaluation undertaken in US K-12 schooling. We write this with humility, as a modest missive to

education stakeholders in diverse Latin American contexts. We recognize the US indeed has a great deal to learn from the tremendous successes of various Latin American educational jurisdictions, as well as the concomitant challenges faced by policy makers in these regions. These are challenging times for policy makers, and the question of how we use evaluation and assessment in our schools is fundamental to our vision and philosophy of education. The questions surrounding testing are never superficial since they are linked to more pressing philosophical questions that probe deeply into the very nature of knowledge. For instance, how can we construct viable explanatory systems of classification, interpretation, and evaluation from the underlying phenomenological descriptions of lived experiences of students, whether through logical inferences or critical/dialectical theories? This is a major task that confronts not only test-makers but education stakeholders at all levels. We must not simply ally ourselves with compatible ideological interpretations about testing but must be willing to challenge all of our fraudulent assumptions. We must risk new understandings that run contrary to our entrenched opinions. The politics of high-stakes testing must remain open to disinterested reflection. This involves a de-dogmatization about the contemporary merits of high-stakes testing and their fundamental necessity. We need to be able to decondition the beliefs and assumptions of our working epistemology; to de-reify and de-automatize everyday reality; and reconsider conditioned attributes that clutter our commonplace observations about measuring and assessing what we deem most pertinent to the process of learning. Without this epistemic decolonization as our guide, we are doomed to repeat and re-entrench the worst elements of capitalism in schooling; a back breaking load to be borne most painfully by those already at the margins of educational power and success. €

- 1 Melissa Lazarín, *Testing Overload in America's Schools* (Washington, DC: Center for American Progress, 2014).
- 2 Arlo Kempf, *The Pedagogy of Standardized Testing: The Radical Effects of Standardization in US and Canadian Education*, (New York: Palgrave, 2016), 15-17.
- 3 Howard Nelson, *Testing More, Teaching Less, What America's Obsession with Student Testing Costs in Money and Lost Instructional Time* (Washington, DC: American Federation of Teachers, 2013), 6-7.



- 3 Arlo Kempf, *The Pedagogy of Standardized Testing: The Radical Effects of Standardization in US and Canadian Education*, (New York: Palgrave, 2016), 15-17. Matthew Lynch, "Education Officials to Re-examine Standardized Testing," *Education Week*, October 22, 2014, accessed October 23, 2014, <http://goo.gl/s6czCW>
- 4 Lizette Alvarez, "States Listen as Parents Give Rampant Testing an F," *New York Times*, November 9, 2014, accessed November 10, 2014, <http://goo.gl/QeLhYd>.
- 5 Howard Nelson, *Testing More, Teaching Less, What America's Obsession with Student Testing Costs in Money and Lost Instructional Time* (Washington, DC: American Federation of Teachers, 2013), 6-7.
- 6 Organization for Economic Cooperation and Development, *Reviews of National Policies for Education: Improving Lower Secondary Schools in Norway* (Paris: OECD, 2011), 9. Arlo Kempf, *The Pedagogy of Standardized Testing: The Radical Effects of Standardization in US and Canadian Education*, (New York: Palgrave, 2016), 15-17
- 7 Organization for Economic Cooperation and Development. *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices Volume 4* (Paris: OECD Publishing, 2013), 147. Arlo Kempf, *The Pedagogy of Standardized Testing: The Radical Effects of Standardization in US and Canadian Education*, (New York: Palgrave, 2016), 15-17
- 8 Organization for Economic Cooperation and Development. *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices Volume 4* (Paris: OECD Publishing, 2013), 147.
- 9 Alfie Kohn, "Standardized Testing and its Victims," *Education Week*, September 27, 2000, accessed October 1, 2014, <http://goo.gl/7ejJKN>.
- 10 Arlo Kempf, *The Pedagogy of Standardized Testing: The Radical Effects of Standardization in US and Canadian Education*, (New York: Palgrave, 2016), 15-17.
- 11 See: Chicagoland Researchers and Advocates for Transformative Education. *Research Brief 1: Testing Today in Context: History, Impact, and Alternatives* (Chicago: CReATE, 2012), 4, accessed January 10, 2014, <http://www.createchicago.org/>; Pasi Sahlberg, *Finnish Lessons: What Can the World Learn from Educational Change in Finland?* (New York, NY: Teachers College Press, 2011); Richard Thorndike, *Measurement and Evaluation in Psychology and Education* (8<sup>th</sup> ed.) (Upper Saddle River, NJ: Prentice Hall, 2011); Howard Wainer, *Uneducated Guesses: Using Evidence to Uncover Misguided education Policies* (Princeton, NJ: Princeton University Press, 2011); James Popham, *Classroom Assessment: What Teachers Need to Know* (Boston, MA: Pearson, 2011); James Popham, "Instructional Insensitivity of Tests: Accountability's Dire Drawback," *Phi Delta Kappan* 89 (2007):146-155; James Popham, *The Truth About Testing: An Educator's Call to Action* (Alexandria, VA: ASCD, 2001); Linda Darling-Hammond, *The Flat World and Education: How America's Commitment to Equity will Determine our Future* (New York, NY: Teachers College Press, 2010); Alfie Kohn, *What Does it Mean to be Well Educated?* (Boston, MA: Beacon Press, 2004); Alfie Kohn, "Standardized Testing Separating Wheat Children from Chaff Children," *AlfieKohn.Org*, 2002, accessed May 3, 2013, <http://goo.gl/AwhS6w>; Peter Sacks, *Standardized Minds: The High Price of America's Testing Culture and What we can do to Change it* (Cambridge, MA: Perseus, 1999); and others.
- 12 Chicagoland Researchers and Advocates for Transformative Education. *Research Brief 1: Testing Today in Context: History, Impact, and Alternatives* (Chicago: CReATE, 2012), 4, accessed January 10, 2014, <http://www.createchicago.org/>.
- 13 William Popham, *Classroom Assessment: What Teachers Need to Know* (Boston, MA: Pearson, 2011), 83.
- 14 Arlo Kempf, *The Pedagogy of Standardized Testing: The Radical Effects of Standardization in US and Canadian Education*, (New York: Palgrave, 2016), 18.
- 15 See: Arlo Kempf, *The Pedagogy of Standardized Testing: The Radical Effects of Standardization in US and Canadian Education*, (New York: Palgrave, 2016), 15-17.; James Popham, *Classroom Assessment: What Teachers Need to Know* (Boston, MA: Pearson, 2011); James Popham, *The Truth about Testing: An Educator's Call to Action* (Alexandria, VA: ASCD, 2001); Richard Murnane and John P. Papay, "Teachers' Views on No Child Left Behind: Support for the Principles, Concerns about the Practices," *Journal of Economic Perspectives* 24, no. 3 (2010): 151-66; Wayne Au, *Unequal by Design: High-Stakes Testing and the Standardization of Inequality* (New York: Routledge, 2007); Linda Darling-Hammond, "Race, Inequality and Educational Accountability: The Irony of 'No Child Left Behind,'" *Race Ethnicity and Education* 10 (2007): 245-260; David Hursh, "Exacerbating Inequality: The Failed Promise of the No Child Left Behind Act." *Race Ethnicity and Education* 10, no. 3 (2007): 295-308; Kate Menken, "Teaching to the Test: How No Child Left Behind Impacts Language Policy, Curriculum, and Instruction for English Language Learners," *Bilingual Research Journal* 30, no. 2 (2006): 521-546; Rod Paige, "No Child Left Behind: The Ongoing Movement for Public Education Reform." *Harvard Educational Review* 76, no. 4 (2006): 461-473; John Rogers, "Forces of Accountability? The Power of Poor Parents in No Child Left Behind," *Harvard Educational Review* 76 (2006): 611-641; Jennifer Booher-Jennings, "Below the Bubble: 'Educational Triage' and the Texas Accountability System," *American Educational Research Journal* 42 (2005): 231-268; Frances Contreras, "Access, Achievement, and Social

- Capital: Standardized Exams and the Latino College-Bound Population" *Journal of Hispanic Higher Education* 4 (2005): 197–214; Jo Boaler, "When Learning no Longer Matters—Standardized Testing and the Creation of Inequality," *Phi Delta Kappan* 84 (2003): 502–506; Mary Barksdale-Ladd, and Karen Thomas, "What's at Stake in High-Stakes Testing: Teachers and Parents Speak Out," *Journal of Teacher Education* 51 (2000): 384–397; and others.
- 16 Provasnik, Stephen, et al., "Highlights From TIMSS 2011: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context," National Center for Education Statistics, 2012, accessed May 2, 2014, [www.nces.ed.gov/TIMSS](http://www.nces.ed.gov/TIMSS).
- 17 National Center for Education Statistics (2012), *The Nation's Report Card: Trends in Academic Progress 2012*, accessed October 21, 2014, <http://nces.ed.gov/pubsearch/pubsinfo.asp>.
- 18 National Center for Education Statistics (2012), *The Nation's Report Card: Trends in Academic Progress 2012*, accessed October 21, 2014, <http://nces.ed.gov/pubsearch/pubsinfo.asp>.
- 19 Arlo Kempf, *The Pedagogy of Standardized Testing: The Radical Effects of Standardization in US and Canadian Education*, (New York: Palgrave, 2016), 14–20.
- 20 Arlo Kempf, *The Pedagogy of Standardized Testing: The Radical Effects of Standardization in US and Canadian Education*, (New York: Palgrave, 2016), 14–20.
- 21 Pasi Sahlberg, "Global Educational Reform Movement is Here!" (PasiSahlberg.com, N.D.), accessed September 6, 2014, <http://goo.gl/52GsWm>. Richard Wilkinson and Kate Pickett. *The Spirit Level: Why More Equal Societies Almost Always Do Better* (London: Allen Lane, 2009).
- 25 Richard Wilkinson and Kate Pickett, *The Spirit Level: Why Equality is Better for Everyone* (London: Penguin, 2010).
- 26 Pasi Sahlberg, "What Canada can Learn from Finland" (2014 R.W. Jackson Lecture, Toronto, Ontario, April 17, 2014). Webcast available at <https://webcasts.welcome2theshow.com/OISE/2230>.
- 27 Andy Hargreaves and Dennis Shirley, *The Fourth Way: The Inspiring Future for Educational Change* (Thousand Oaks, CA: Corwin Press, 2009), 62.
- 28 Arlo Kempf, *The Pedagogy of Standardized Testing: The Radical Effects of Standardization in US and Canadian Education*, (New York: Palgrave, 2016), 14–20.
- 29 John Marsh, *Class Dismissed? Why We Cannot Teach or Learn Our Way Out of Equality* (New York: Monthly Review Press, 2011), 89.
- 30 John Marsh, *Class Dismissed? Why We Cannot Teach or Learn Our Way Out of Equality* (New York: Monthly Review Press, 2011), 86.
- 31 Norman Doidge, *The Brain That Changes Itself: Stories of Personal Triumph from the Frontiers of Brain Science* (New York: Viking, 2007), 309–10.
- 32 John Marsh, *Class Dismissed? Why We Cannot Teach or Learn Our Way Out of Equality* (New York: Monthly Review Press, 2011), 57.
- 33 John Marsh, *Class Dismissed? Why We Cannot Teach or Learn Our Way Out of Equality* (New York: Monthly Review Press, 2011), 182.
- 34 John Marsh, *Class Dismissed? Why We Cannot Teach or Learn Our Way Out of Equality* (New York: Monthly Review Press, 2011), 210.

## VOICES FROM THE CONFERENCE

### Education-policy decisions: the link between evidence and values

"Values open up new horizons and public and human aspirations, while evidence can point to both obstacles and possibilities", says the author. "Evidence can support precise, correct diagnoses of situations or problems, but values provide the social leverage needed to guide interventions aimed at addressing the said problems or situations and justify changes that are both possible and necessary", he assures.

FRANCISCO MIRANDA LÓPEZ  
Head of the Educational Policy  
and Regulation Unit of the INEE  
[fmiranda@inee.edu.mx](mailto:fmiranda@inee.edu.mx)

Decisions are taken in different spheres and at different levels, and involve different hierarchies. How one defines what the decision-making process is, or should be, depends on the frame of reference used and the role assigned to knowledge, values or analysis within the said frame.

If one believes that knowledge or reasoning are involved in the said process, then data will play a fundamental role in ascertaining what view of reality informs decision-making. However, since this very belief in the rationality and "value" of evidence would also appear to assume an encounter with values, then, in order to analyze decisions, we must understand how they take shape and how they put facts and values into practice, as well as how convictions, ideas and interests interact with facts and realities, and vice versa (Parsons, 2013).

Assuredly recent discussions about public-policy analysis have been pervaded by the growing conviction that policy-implementation decisions must be based on evidence. An important influence in this regard is the so-called evidence-based-policy (EBP) approach, whereby public-policy decisions and choices



among public-policy options must be based on the overt, purposeful use of the best available evidence stemming from facts, data and relationships rooted in empirical conclusions that are not founded on the subjective views of a small number of protagonists, but rather on objective measurement using relevant, reliable methods and techniques (Davis, 2000; Aguilar, 2005; Bracho, 2010).

Although the above idea is very convincing and has much to recommend it, it does not completely take stock of the role that other subjective, judgmental factors play in the decision-making process. While EBP helps to ensure that decisions are more rational and objective, guarding against decisions that are whimsical, doctrinal, irrational or headstrong, it does not take into account the fact that all policy-making decisions involve value judgments - i.e. ultimately it does not envisage other logics, based on principles and values, that also inform and justify the decisions that are made and can play an important part in mediating and leveraging the data, and the models stemming therefrom, for the purpose of making decisions about public policy (Simon, 1984; Majone, 1989; Bracho, 2011).

What we wish to stress, in this brief article, is that EBP logic needs to serve as a basis for making rational, value-related decisions, rather than being used to support decisions that only contemplate the desired ends, since we are convinced that this will enable us to make the data more meaningful and orient the way they are used to solve problems, provide input to debates, and inform policy-making, which, in modern “public-policy-governed countries”, needs, among other things, to involve citizens, respect and further human rights, solve social problems, and foster democratic government, transparency and accountability, all of which are basic principles adherence to which is all the more essential in government policy-making.

With regard to the aforesaid imperatives and requisites, we go on below to provide a brief breakdown of the different meanings assigned to evidence and values, and of their felicities and shortfalls.

#### **Evidence-based reflection**

It can be argued that the notion of evidence rests on at least three core concepts that support its claims to be valid for purposes of decision-making. The first of these is the concept of “hard evidence” that supports decision-making, as against “soft concepts”

that lead to subjective, arbitrary, biased decision-making; the next is the concept of the rigorously constructed “evidence-based approach” as a means of avoiding simplistic, mechanical judgments; and the third is the concept of “evidence as an objective affirmation” that stresses the measurement of problems as a prerequisite for sizing them up, analyzing them and thereby coming up with possible solutions.

Evidence deemed to be “hard” is assumed to consist of relevant information that does not arise from people’s subjectivity – i.e. information that is not distorted by perceptions or opinions based on sympathy, hearsay or empathy vis-à-vis a given fact or situation. What distinguishes hard data is that they exist independently of people and, although they may give rise to the latter’s actions or behaviors, constitute an external, objective reality that entails its own thought processes and behavior (Hitch and McKean, 1962; Alain and Smith, 1971; Quade, 1982; Corzo, 2013). Hence, it is considered that, when a decision is based on the best available evidence, it is more likely to solve the problem that it is seeking to solve or address, so long as it is not the butt of personal judgements, disagreements between people or groups, or the opinions of those who play a part in formulating it, implementing it, or analyzing its results.

On the other hand, evidence is commonly associated with the procedures or methods that are used to produce it, which is not synonymous with “hard data”, since the latter are found, while evidence is always built, rather than floating around waiting to be caught, being the intentional result of a method that involves various conceptual resources, disciplinary frameworks, technical resources and operational contexts that are different from -and in many ways more complex than- those required for data analysis (Aguilar, 2005; Muñoz y González, 2010). Hence, the method renders the evidence valid by enabling it to be replicated and compared, thus depending on who proposes or propounds it, rather than on the different mental and technical operations that underprop it and can render it “falsifiable” via different testing or refutation mechanisms.

Finally, strongly associated with the concepts of “hard data” and “hard method”, we use the term “magnitude” to refer to evidence. One of the most characteristic features of all hard evidence is that it results

from the use of numerical scales to measure and observe aspects and facets of reality. Hence, by establishing objective magnitudes, we can more accurately measure and define the problem, breaking it down into its different components and rendering it open to analysis whereby it can be mathematically or logically duplicated so as to delve deeper into it, create scenarios for experimentation and the establishment of predictability, or associate with other variables to ascertain how “flexibly” it behaves. Thus, how sensible and efficient decisions are depends on the extent to which they are based on the measurement of the problems analyzed, the precision with which their magnitudes are established, and their potential variability, all of which, together, enable us to plan precise interventions aimed at solving, diminishing, limiting or redefining them (Aguilar, 2005; Muñoz and González, 2010), and also to establish criteria and mechanisms to measure the success of the said interventions.

#### **Proposals based on values and principles**

Just as our everyday concept of what is good and bad, correct or incorrect, fair or unfair influences our decisions and actions, in the same way the ideas and conceptions of society and politicians about certain values influence the orientations of the policy decisions that are made, functioning as a kind of router in our assessment of the reality we perceive. With varying intensity, these routers serve, at least initially, to channel or generate the meaning of everything we normally experience in our private lives, in the public domain, and in political or institutional life.

In the vast literature about research into values, one can identify at least three core concepts that bear keeping in mind in the context of decision-making and public-policy processes. First, values are interpreted as principles, becoming ethical mainstays, ontological bases and compasses that guide our behavior, second, they are like joints or interfaces between collective, political and private or small-group life, and third they serve as touchstones for judging the actions and behavior of our fellow men.

As principles, values delimit meaning, and hence govern people’s behavior, also speaking to individuals and binding them via their belief in, or acceptance of, the con-

tents or attitudes that they prescribe or determine (Aguilar, 2007; Weimer and Vining, 2010; Bracho, 2011; Arellano and Blanco, 2013, Merino, 2013). In this way values have to do with the foundations, origins, main tenets or *raison d'être* of things, thus forging strong links among individuals and groups that have -or seek to create- common factors that identify and unify them and are recognizable in all physical, cultural, social or historical contexts.

However, values serve not only as principles in the abovementioned way, but also as social levers or instruments that enable communication between -and containment and development of- individuals and society, a role that they play when they push and guide individual action towards more meaningful social spheres such as groups, communities, countries or states (Schwartz, 1999; Lindblom, 1992; Asthama, Richardson and Halliday, 2002). Values serve as important mechanisms for reduplicating the social order, either for the purpose of strengthening identification between individuals and larger social groups and affirming the former's membership of the latter, or in order to achieve congruence between sociopolitical criteria and the criteria governing individual or group behavior.

Finally, values are closely linked to the criteria whereby we interpret and evaluate reality and human behavior, and on which we base value judgements and translate them into certain judgments about the state, nature of -or results stemming from- a given thing, event or process pertaining to material, social or human reality. Moreover, by functioning in this way, values serve not only to determine the place we can assign to things within given categories, but may also be open to some kind of inner grading or stratification that enables us to measure the distance between certain features or traits of the aforesaid things and certain desired values.

### Linkage in evaluation policy

As is clear from the above comments and concepts, initial conclusions regarding grounds for decisions -above all political decisions- must be based on a coordinated effort to reach decisions that are both effective and legitimate. In the final analysis we are talking about decision-making that is unified and structured in line with needs and contexts where reasoning based on ends and reasoning based on values come together in

strategies aimed at serving the public good. In other words, policies and the decisions that stem from them are the result of a tense, composite process that potentially balances the gathering and use of evidence and the creation of meaning and structural content oriented by fundamental principles or values.

While evidence provides empirical support and underpins the linkage between ends and means that is part and parcel of all policy decisions, values can increase the amount of knowledge that the said evidence can contribute by demanding more data and measurements regarding the basic fulfillment of moral, legal or human obligations. However, the opposite may also be the case, since, as result of data regarding viability and its technical and scientific underpinning, or based on the making of technical or historical comparisons during political debate, evidence can dampen the ideological or political zeal of some decision-makers to address social problems and make the necessary means available.

Values open up new horizons and public and human aspirations, while evidence can point to both obstacles and possibilities. Evidence can help us to determine what is politically feasible, but values and principles can lead to concerted political efforts to overcome limitations and open up new worlds of possibility. Evidence can support precise, correct diagnoses of situations or problems, but values provide the social leverage needed to guide interventions aimed at addressing the said problems or situations and justify changes that are both possible and necessary.

In short, the relationship between evidence and government policy can by no means be linear or automatic. As Eugene Bardach asserts (1980), in politics no methodology is universally valid, and so, rather than seeking methodological anarchism, one should, to the extent that the context allows, move towards plurality, which implies various interventions that are not just technical, but also involve different ethical, political and social criteria. The quest for fundamental balanced decisions that are technically sound on the one hand, and socially acceptable, fiscally responsible and institutionally transparent on the other hand, is probably one of the most sensitive aspects of public-policy formulation, added to which such decisions need to be participative, take into account the points of view

of those involved or benefited, and also meet the legitimacy criteria that underpin democratic governments, as well as satisfying transparency and accountability criteria.

This is especially pertinent for evaluation policy, an area in which one seeks not only to make effective, legitimate decisions that render evaluation relevant, fair, adequate and transparent in order to ascertain the state or condition of education, but also to support education-policy decisions.

The aim of evaluation policy is both to remedy problems and fill gaps in evaluation, and also to foster improvement. In other words, one seeks, on the one hand, to create a connection between evidence and values that makes it possible to ascertain the state of the components, processes and results of the Mexican Education System, and, on the other hand, to build logical, empirical, value-based and policy links between the status quo and the proposed policy action, which entails going beyond the strict terrain of evaluation and entering the domain of planning, government administration and policy-making.

The processes whereby evidence and values are linked or separated in education policy would seem to imply similar processes in government policy-making. Even when evidence and the weight thereof are the same, or government action adheres to the same value benchmarks in both evaluation policy and education policy, in fact there will always be margins of linkage or separation that are sure to determine the effectiveness, credibility and, ultimately, the direction of government action aimed at improving education, with regard both to quality and to equity.

In the worst case, the aforesaid dynamic could lead to tensions and differences in the building of consensus and the positing of basic values, not to mention the problems associated with issues of legitimacy (i.e. that of data, methodology and measurability) vis-à-vis an educational problem and the public debate to which the latter can give rise. €

### References

- Aguilar, L. F. (2005). Introductory study, in Majone, G. *Evidencia, argumentación y persuasión de la formulación de políticas*. Mexico: CNCYPAP/FCE.
- Aguilar, L. F. (2007). *El estudio de las políticas públicas*. Mexico: Porrúa.
- Alain C. E. and Smith, W. (1971). *How much is enough*. New York: Harper and Row.

- Arellano, D. and Blanco, F. (2013). *Políticas públicas y democracia*. Mexico: Instituto Federal Electoral.
- Asthama, S., Richardson, S., and Halliday, J. (2002). Partnership working in public policy provision. *Social Policy and Administration*, Vol. 36, No.7, pp. 780.
- Bardach, E. (1980). *The Implementation Game*. Cambridge: The MIT Press.
- Bracho, T. (2010). Políticas públicas basadas en la evidencia. La política pública como acción informada y objeto de investigación, en Merino, M. and Cejudo, G. (Compilers). *Problemas, decisiones, soluciones. Enfoques de política pública*. Mexico: FCE/CIDE.
- Bracho, T. (2011). El papel de los valores en la evaluación de políticas para sistemas educativos complejos. *Revista Mexicana de Investigación Educativa*, 16 (50).
- Corzo, F. (2013). *Diseño de políticas públicas*. Mexico: Grupo Editorial y de Investigación
- Polaris, S. A. de C. V.
- Corzo, F. (2013). *Diseño de políticas públicas*. Mexico: Grupo Editorial y de Investigación Polaris S. A. de C. V.
- Davies, P. (2000). The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education*, 26 (3 y 4), pp. 365-378.
- Hitch, C. J. and McKean, R. (1962). *The Economics of Defense in the Nuclear Age*. Cambridge: Harvard University Press.
- Lindblom, Ch. (1992). *El proceso de elaboración de políticas públicas*. Madrid: Ministerio para las Administraciones Públicas.
- Majone, G. (1989). *Evidence, Argument and Persuasion in the Policy Process*. Yale University Press.
- Merino, M. (2013). *Políticas públicas. Ensayo sobre la intervención del Estado en la solución de problemas públicos*. Mexico: CIDE.
- Muñoz, M. and González, J. (2010). *Valoración de la calidad de la evidencia y fuerza de las recomendaciones*. Barcelona: Sociedad Española de Medicina Familiar y Comunitaria.
- Parsons, D. W. *Public Policy: An Introduction to the Theory and Practice of Policy Analysis*, Elgar, Edward Publishing, Inc. (1995). Published in Spanish by Flacso, Mexico. (2007)
- Quade, E. S. (1982). *Analysis (or Public Decisions)*. Amsterdam, North-Holland.
- Schwartz, S. (1999). A theory of cultural values and some implications for work. *Applied Psychology*, 48 (1), pp. 23-47.
- Simon, H. (1984). *La nueva ciencia de la decisión gerencial*. Buenos Aires: El Ateneo.
- Weimer, D. and Vining, A. (2010). *Policy Analysis: Concepts and Practice*. USA: Routledge.

## VOICES FROM THE CONFERENCE FACING OUR CHALLENGES

### Challenges, aims and work: decisions and educational evaluation in Querétaro

In the interview transcribed below, which doesn't just focus on education, but also covers policy and syllabus design and the use of information in the light of the recent results achieved via evaluation, **José Alfredo Botello Montes**, Minister of Education for the State of Querétaro, talks with the Gazette about his states experiences with educational evaluation and the challenges that it faces.

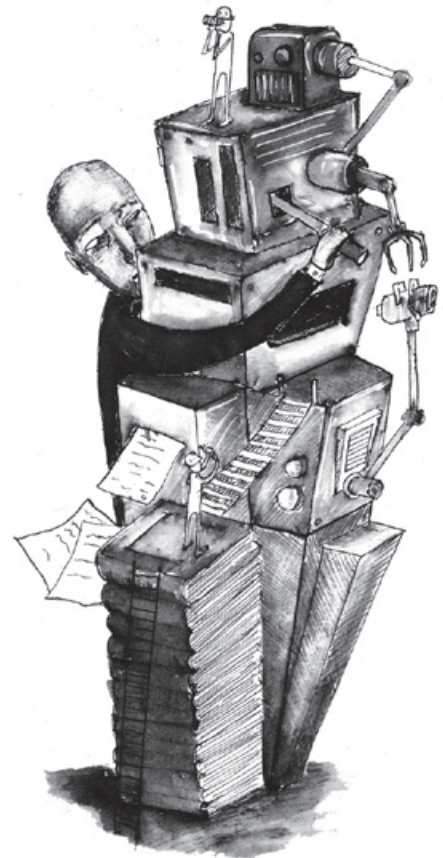
**M**r. Botello Montes, who is Minister of Education for the State of Querétaro, points out that the strategic aims of the ministry he heads include:

**I: raising educational quality** by improving school conditions so that students -whether children, youths or adults- can acquire the knowledge, and develop the skills, values and attitudes, that they need in order to have a rich personal, family and social, be responsible, committee citizens, and go on learning throughout their lives.

**II: ensuring equality** in terms of access to, continuation in, and completion of, all the types, modalities and levels of education, with priority being afforded to the most disadvantaged students, groups and locations.

**III: expanding coverage** by increasing, and making more flexible, the opportunity for people to enter, remain in, all the types, modalities and levels of education, ensuring that they have access to cultural services and resources, sports and recreational activities, and scientific and technical amenities and events.

**IV: Improving institutional management** by consolidating the running and linkage of the state-level education system in order to ensure that education policies are effective continually evaluated, that resources are used efficiently, transparently and account-





ably, that the needs of both students and society are satisfied, and that management centers on schools and classrooms.

Asserting that the position he holds affords him a big opportunity to serve others in his state, which is an important region of Mexico, Botello talks about the main input that the state of Querétaro has managed to obtain via educational evaluation and how the said input has been used in order to design the current policies and programs:

“Of course,” he says, “the evaluation currently being carried out by the National Educational Evaluation Institute (Spanish acronym: INEE) is very important, since it provides us with results that have an impact at both the national and international levels, mainly in the areas of Language and Communication and Mathematical Reasoning (See Table 1 and Figure 1)”.

“For its part”, remarks Botello, “Planea has become a very important means of measuring the results of the local education system, due to its emphasis on the acquisition of basic learning and curriculum-based competencies. After the Education Reform that began in our country a couple of years ago, it is essential that we obtain precise information about the suitability for the teaching profession of our teacher trainees and in-service teachers based on their performance in the said evaluations”.

“Furthermore, Planea helps us to ascertain our students’ progress and find out our state-level education systems areas of opportunity for ongoing improvement, as well to measure our performance at the national and international levels”.

Regarding the aforesaid areas of opportunity and the design of policies and programs aimed at exploiting them, Botello, a former federal congressman, comments:

“We need to work with the teachers, the managers and the students themselves to hone their skills in the important areas of language, communication and mathematical reasoning”.

“Though we’ve already addressed the latter area -i.e. including it in the curriculum- we haven’t given it the importance it merits, notwithstanding the fact that it’s the basis for all logical thought, which is why we’re offering the largest number of workshops, and why one of the policies we’re going to implement this year focuses heavily on language and mathematical reasoning”.

**Table 1. Planea 2015 results for the state of Querétaro**

		Placing nationwide	% of students placing in levels II to IV	Countrywide average
6 <sup>th</sup> year of Primary School	Language and Communication	9 <sup>th</sup>	54.6%	50.4%
	Mathematics	8 <sup>th</sup>	43.2%	39.5%
3 <sup>rd</sup> year of Secondary School.	Language and Communication	4 <sup>th</sup>	76.3%	70.5%
	Mathematics	3 <sup>rd</sup>	42.5%	34.6%

Note: Pertains to the National Pan for the Evaluation of Learning Outcomes Examination (Spanish acronym: Planea) elementary-level evaluation in Mathematics and Language and Communication that was administered in June, 2105, along with a questionnaire about coexistence skills, to 53,220 students in the final primary and secondary grades in 1,833 schools in Querétaro. Source: Ministry of Education of the State of Querétaro.

**Figure 1: 2015 Planea upper-secondary-level results for Querétaro**

The results of the 2015 Planea upper-secondary-level examination were as follows:

42.2% of the students in Querétaro placed in the highest performance levels (II and IV) in Language and Communication (i.e. reading comprehension), compared with 36% nationwide, the best results, in descending order, being obtained in the states of Puebla (1<sup>st</sup> place), Baja California (2<sup>nd</sup> place), Durango (3<sup>rd</sup> place), Jalisco (4<sup>th</sup> place) and Querétaro (5<sup>th</sup> place).

22.2% of the students in Querétaro placed in the highest performance levels in Mathematics, compared with an 18.8% nationwide average), the best results, in descending order, being obtained in the states of Durango (1<sup>st</sup> place), Puebla (2<sup>nd</sup> place), Sonora (3<sup>rd</sup> place), Sinaloa (4<sup>th</sup> place), Baja California (5<sup>th</sup> place), Querétaro (6<sup>th</sup> place) and Chihuahua (7<sup>th</sup> place).

Source: Ministry of Education of the State of Querétaro

“Communication, for its part, is an area of opportunity stemming from the Educational Reform. We have the School Technical Committees (STCs), in which parents, teachers and the education authorities work together in order to proactively promote ongoing improvement in schools, and these also serve to help school principals and teachers to monitor the educational community’s endeavors to implement the poli-

cies that are needed in order to make such improvements (See Figure 2)”.

“We should take advantage not only of the instruments that are currently available, such as the School and Zonal Technical Committees, but also of the direct links that should exist among school principals, school management, teachers and parents. Having noticed that the latter are not involved, on an ongoing or regular basis, in what goes

**Figure 2: Results of the 2015 competitive process for appointing new elementary-level teachers.**

From the 11th to 19<sup>th</sup> of July, 2,341 of the 2,729 candidates who had applied to participate in the selection process for appointment to elementary-level teaching posts in the State of Querétaro (i.e. 85.78% of the total registrées), with the following results:

69.76% of the examinees (i.e. 1,633 out of 2,341 candidates) got a “Suitable” grade, putting our state in first place countrywide, with the biggest percentage of “Suitable” grades, as against a national average of 51.91%.

30.24% of the candidates (707 people) were awarded an “Unsuitable” grade, as compared with a national average of 48.09%.

16.15% of the candidates who were graded “Suitable” placed in the Very High Performance group, 24.86% in the High Performance group, 27.72% in the Good Performance group, and just 1.03% in the Satisfactory Performance group.

72.88% of the teacher-training-college graduates who took the examination got a “Suitable” grade, making Querétaro the state with the fourth highest number of such graduates with “Suitable” grades countrywide.

Just 26.01% of all the teacher-training-college graduates who took the examination in our state -i.e. one of the lowest percentages in all of Mexico’s 32 states- received an “Unsuitable” grade.

Source: Ministry of Education of the State of Querétaro

**Table 2. Coverage in the Querétaro Education System**

Level	Coverage: new age ranges <sup>1</sup>			Coverage: traditional age ranges <sup>2</sup>		
	Registered	Population*	%	Registered	Population*	%
Pre-school	85 065	112 116	75.9	85 065	112 116	75.9
Primary	248 608	227 087	109.5	248 608	264 884	93.9
Lower-sec- ondary	120 780	113 458	106.5	120 780	113 548	106.4
Upper-sec- ondary	75 932	113 568	66.9	75 932	113 320	67.0
Lower-profes- sional	219	113 568	0.2	219	113 320	0.2
Senior-high	75 713	113 568	66.7	75 713	113 320	66.8
Higher**	65 790	185 466	35.5	65 790	183 815	35.8

Source: Data: start of the 2014-2015 school-year. Statistical Summary. Data: start of the 2014-2015 school-year. Education Portal. Government of the State of Querétaro.

\* Population projections: National Population Council (Spanish acronym: Conapo, April 2013 / \*\* Bachelor’s Degree and Classroom-taught Advanced Technical Diploma.

on inside our schools, I believe that another area of opportunity exists to find ways to involve them more in their children’s education and prevent them from always being on the sidelines, and I’ve noted that our teachers are keen to help in this endeavor”.

“Additionally, given the size of our state’s education system, we face the problem of turnover in our schools due to retirement (See Table 2). It’s hard to find enough class teachers and one of our challenges is to issue calls for candidates to take the examinations for admission to the National Teaching in order to cover vacant positions with suitable staff”.

Assuredly, the dissemination of educational-evaluation results is of the essence, and on this topic Botello comments:

“Though I only recently took up my post in October of 2015, one of my aims is to ensure that our School Technical Committees are duly set up and that both my own Ministry and the Federal Ministry of Public Education can provide them with feedback on evaluation results, though we also want to look at the issue of ensuring participation by the members of the said committees, which we believe to be a big challenge not only locally, but also countrywide”.

A former teacher at the Monterrey Technical and Higher-studies Institute (Spanish acronym: ITESM), Botello Montes stresses his state’s strengths:

“The greatest strength I’ve found are the very professional teaching and managerial staff, who’ve worked their way up through the professional-teaching system. I’ve even met people with Ph.D’s and master’s degrees in education who are class teachers or school principals. Our staff are very well trained and enthusiastic”.

“In our Department of Elementary Education we encountered three different planning areas which would assume responsibility for evaluation-related tasks. There was a primary section, a pre-school one and a secondary one, each operating separately without any links to the others. We’re currently reengineering their management so as to fuse them into a single area and ensure very good communication among them, so as to achieve a new, uniform evaluation system that serves all the different types of education, rather than a jigsaw puzzle. We want to work as a team, a body whose arms, legs and head work in unison”.

On the subject of the National Educational Evaluation Policy and the Dialogues

with the Educational Authorities aimed at formulating it, he says:

“The National Educational Evaluation System Conference is an excellent mechanism implemented by the INEE. I was able to attend the meeting that was held in October, and found that it was a very good forum for giving input and exchanging opinions. They showed us data indicating which actions need to be taken and provided us with feedback as a basis for decision-making”.

“I believe that participation in the said Dialogues should not be limited to the Ministry of Education, but also involve the staff

who are directly involved in evaluation, and I also think that we should give feedback to the rest of the education authorities on our views about the application of the evaluations. In general, I find these meetings, where we can meet and understand each other, and exchange opinions, to be a very useful way of ensuring that state-level policy doesn't become divorced from federal policy”.

“Our ultimate aim is to produce good, well trained, well educated citizens. This is the big task facing our country, its states and also, of course, its municipalities”. €

- 1 New age ranges: preschool: ages 3, 4 and 5; ages 6 to 11; lower-secondary: ages 12 to 14; upper-secondary: ages 15 to 17; higher: ages 18 to 22.
- 2 Traditional age ranges: preschool: ages 3, 4 and 5; primary: ages 6 to 12; lower-secondary: ages 13 to 15; upper-secondary: ages 16 to 18; higher: ages 19 to 23.



Visit the website of the Ministry of Education of the State of Querétaro at: <http://www.queretaro.gob.mx/educacion/>

## VOICES FROM THE CONFERENCE FACING OUR CHALLENGES

### An overview of education and evaluation in Sonora: how to overcome the challenges

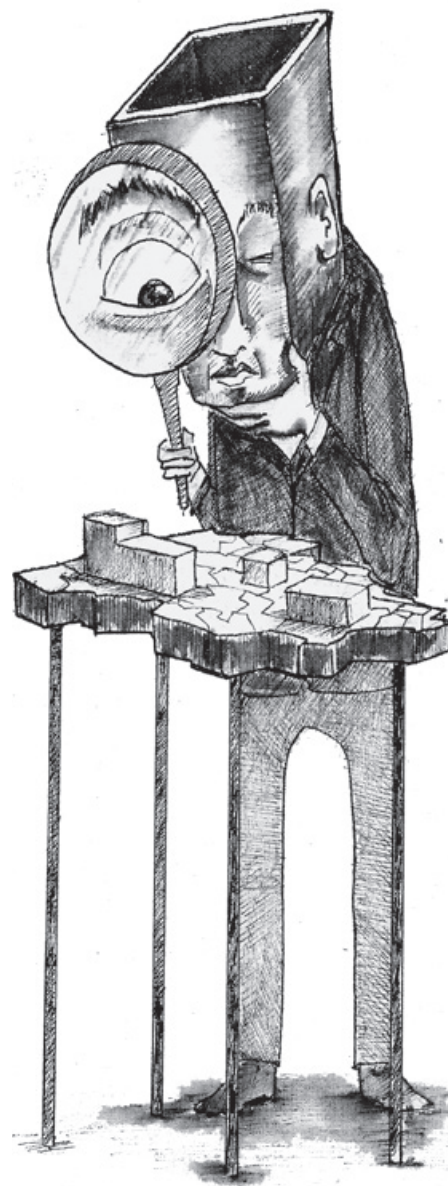
“To raise quality, we need an evaluation system that produces information that's relevant for decision-making”, says **Ernesto De Lucas Hopkins**, Minister of Education and Culture for the State of Sonora, in the interview transcribed below, where he talks about the strategies that he is pursuing in order to once more make his state a leader in the field of education.

“For us to create high-quality educational options, the State needs to take actions aimed at achieving renewed levels of efficiency, effectiveness and equity in its policy-making,” states Ernesto De Lucas Hopkins, Minister of Education for the State of Sonora. “In our country, the administration of President Enrique Peña Nieto is carrying out an Educational Reform the main aim of which is to create a high-quality education system for all, as made manifest in the amendment to Article Three of the Mexican Constitution, which was voted into law by all the members of both houses of Congress”.

“Furthermore, with the decentralization that confers autonomous status on the National Institute for Educational Evaluation (Spanish acronym: INEE) the evaluation systems charged with fostering high-quality in education have been rendered legal and institutionalized. In this regard, while not underestimating the challenges faced by Sonora with regard to coverage and satisfaction of the existing demand -above all at the upper-secondary and higher-studies levels- the government of our state, headed by our governor, Claudia Pavlovich Arellano, has made high-quality an across-the-board criterion for the design of educational policies and programs”.

#### Evaluation for improvement

Pavlovich Arellano, who held various government posts in the State of Sonora during the ten years before he took up his present post, asserts: “An evaluation system that produces timely, relevant information for purposes of decision-making, and is accountable both for its results and for the educational processes stemming from them, is a prerequisite for raising the quality of Sonora's education system. The input from evaluation should be used to formulate government policies aimed at raising the quality of education and fostering equal opportunity in education. The main input includes the generation, analysis and dissemination of information and knowledge about education, the design and implementation of a system of benchmarks for ascertaining the quality of education in our state, the carrying out and promotion of research that un-



derpins the design of specific programs in line with the needs of Sonora to strengthen areas such as infrastructure and equipment, professional training and development, and remuneration systems. In the context of the current reform, other very important aims have to do with student learning outcomes and the suitability of teaching and managerial staff”.

“Finally, one of the current state government’s priorities is to include efficiency and effectiveness in the use of public resources as specific input requisites for government-policy design, along with the need for the results to be in harmony with the 2016-2021 State Development Plan and to manifest themselves in the Sectoral Education Plan for the State of Sonora”.

### The challenges of information use

“Now, with regard to the challenges posed by evaluation,” says Pavlovich Arellano, a one-time federal congressman, “it should be stressed that we carried out a deep-reaching, wide-ranging diagnosis of the education sector both by system (input, processes and products) and by type and level of education, likewise evaluating the different policies and programs implemented by the administration that preceded us. However, this was no easy task, since we found either no information and/or information that had been tampered with, which made it hard for us to find a clear, reliable starting point”.

“For the above reason, it’s hard for us in the current administration, and our successors, to come up with timely, accurate information that enables us to take appropriate educational-policy decisions that provide a basis for evaluating our performance, which is why we propose the relaunching of a state-level evaluation system that can work hand-in-hand with the INEE to produce and disseminate input for the design and implantation of actions -above all those needed for our state to exercise the powers assigned to for the purpose of carrying out the Educational Reform – e.g. to suggest complementary profiles, norms and benchmarks, to choose and train evaluators in accordance with the pertinent INEE guidelines, to punctually issue and broadcast public invitations to take part in teacher-selection procedures, to implement the different evaluation processes in accordance with the General Law Governing the Professional Teaching Service (Spanish acronym: LGSPD)

and the guidelines stemming therefrom that are established by the INEE, to disseminate results and recommendations for the design of the of the pertinent ongoing-training and professional-development programs, and to sign such agreements with authorized government institutions as are needed to raise the quality of education, etc.”.

“Outstanding among the members of the previous administration (2009-2015) is De Lucas, who was the national director of ProMéxico (*a trust fund of the Mexican Government, pertaining to the Ministry of Economy, that promotes international trade and investment*) and also of the Institute for Mexicans Abroad (Spanish acronym: IME). We received an education system with serious failings not only in its teacher-training and teacher-development programs, but also in its infrastructure, just to mention the most important ones, and hence the high quality -attested to by various external evaluations- that our state’s education was one renowned for was eroded”.

“On top of the above, the lack of information about the Educational Reform and the dearth of clarity and transparency in the way the LGSPD and its processes were implemented led to a split among our teachers that had a strong impact on our society and impeded the achievement of the culture and employer-employee relationship that were necessary for the signing of agreements and the implementation of strategies, including the dissemination of information and the training of all those involved, to enable the Reform to succeed. As a result, the general unease that prevailed among our teachers delayed the implementation of these things and led to chaos”.

“The lack of transparency in the use of the public resources devoted to education led to the paralysis of educational programs and projects, especially at the upper secondary level. The decentralized subsystems that the state government received were suffering from serious economic problems and, finally, a very worrying matter was the decline in the educational-achievement results of Sonora’s students”.

### The design of the sectoral education program

“As can be seen,” says De Lucas Hopkins, who holds a master’s degree in Law and Government from the Washington College of Law, “given the condition in which we re-

ceived the Sonoran education system at the beginning of this administration, we face big challenges that need to be tackled via government policies based on strategies and lines of action whose priority is to win back the trust of our state’s citizens by involving them in government decision-making and fostering cooperation in all the spheres of public life. These aims are enshrined in the 2016-2021 State Development, which sees the recovery of trust as a challenge that revolves around efficiency – i.e. an administration that responds to the population’s needs via responsible action, the proper handling of public resources, and the implementation of innovative government policies with the three essential features of mainstreaming, equality and competitiveness”.

“The aforesaid should manifest itself in the 2016-2021 Sectoral Education Plan, which will orient our education system, and in which it will be essential for us to have a state-level evaluation system that enables us to make timely decisions and provides feedback about educational programs – i.e. a dynamic, reliable system of benchmarks”.

“It’s clear to us that, in order to once more achieve the high-quality education that we Sonorans’ were accustomed to, we’ll need to need to ensure a minimal level of normalcy in our state’s education system, give more administrative autonomy to our schools, albeit with full state-government support, restart and analyze the training and professional-development programs for teachers and other school staff in the context of the Educational Reform, work with the federal authorities to implement programs for updating, maintaining and building educational infrastructure and supplying equipment and materials, and restart the support and reward programs that made Sonora a leader in educational achievement and enabled it to reduce student lag and dropout rates”.

### Strategies

“Focusing on the aforementioned challenges,” says De Lucas Hopkins, “we’ve based our actions on the following strategies:

- a) The transparent use of the public resources earmarked for education, a focus on information systems that will enable us to win back Sonorans’ trust in their authorities, and the cutting of red tape so as to simplify and reduce paperwork.

- b) The dissemination of information about both the Educational Reform and the secondary laws and processes that form part of it, based on the setting up of the State Office for the Coordination of the Professional Teaching Service (Spanish acronym: CNSPD) and the Office for Liaison Between State Institutions and the Federal Education Authorities and the INEE, the training of teachers so that they can duly and opportunely comply with the different performance-evaluation processes, the encouragement of people to take the evaluation for entry to, and promotion within, the Professional Teaching Service, ensuring that its results are transparent, and the setting up of support programs, including scholarships and free school uniforms, based on the transparent use of resources and the cutting of red tape.
- c) The strengthening of infrastructure, via the updating and maintenance of schools and the provision of materials and furniture.
- d) The setting up of a program to support the analysis of learning-outcome evaluation results, aimed at designing and implementing specific programs for each school and every group, based on mentoring and tutoring for students at risk of lagging behind or dropping out.
- e) The raising of educational quality via a work program to help the schools that got the lowest results in the Planea examination that the INEE has been administering since the 2014-2015 school year”.

“Now high quality and equity,” De Lucas, who was chairman of the Institutional Revolutionary Party (Spanish acronym PRI) State Steering Committee for Sonora in 2008, goes on to say “are the foundations on which our country’s -and hence our state’s- education system should rest”.

“The Sonoran Ministry of Education believes that the monitoring of educational improvement should, as a matter of priority, be accompanied by the other things that enable students to achieve maximal learning outcomes – i.e. teaching-learning materials and methods, organization, infrastructure, equipment, and hard work on the part of teachers and school principals”.

“Since the national policy for raising educational quality is undeniably rooted in

the Reform, it’s up to the state government to ensure the full exercise of the powers and obligations that the LGSPD confers on the local education authority for the purpose of ensuring the efficient, effective running of the education system, and to this end liaison and ongoing communication with the AEF and the INEE are essential if we are to fully comply with the LGSPD and the guidelines that stem from it, or, failing this, propose measures for perfecting and improving education in our state based on the latter’s specific experiences and particularities, without losing sight of the shared aim of strengthening both the local and national education systems and raising their quality”.

“We believe that, in order to raise educational quality in our state, we need, among other things, to:

- a) Save and strengthen the State-of-Sonora Institute for Educational Evaluation (Spanish acronym: IEEES).
- b) Set up a State Office for the Coordination of the Professional Teaching Service.
- c) Formulate government policy based on the results of the Census of Elementary-level and Special-education Schools, Teachers and Students.
- d) As a matter of priority, set up a program for the training and professional development of elementary- and lower-secondary-level teachers and school principals.
- e) Design educational programs based on the results of the external evaluations”.

“Currently, we’re in the process of designing strategies to reverse the Planea results, which constitute a big challenge, since Sonora placed 18<sup>th</sup> in Language and Communication, and next to last in Mathematics at the primary level, and 27<sup>th</sup> in both the aforementioned sections of the examination at the secondary level”.

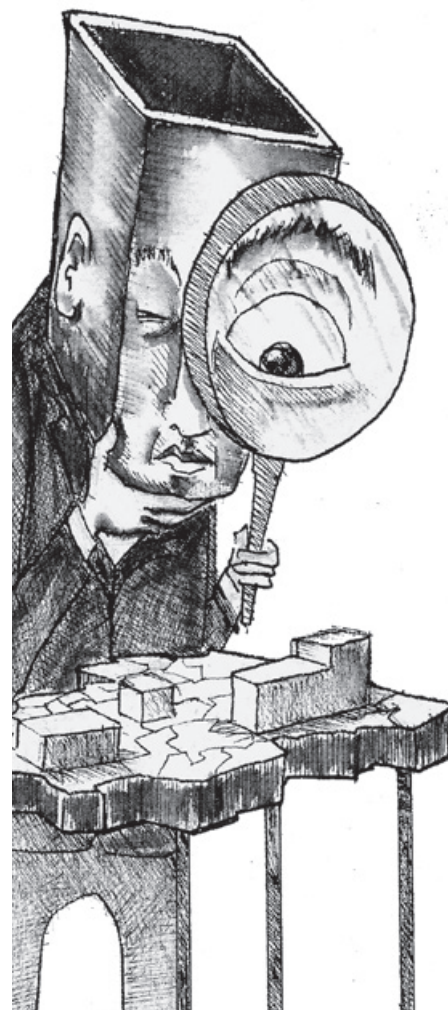
“In order to tackle this challenge, we went about identifying the schools that got the worst results, in order to come up with strategies involving their principals, teachers, students and parents, while also not forgetting to strengthen the schools that got good results. Though the aforesaid strategies are still being implemented, we’re sure that they’ll enable us to get better Planea results, since we have prior experience of implementing support programs, based on

results of the evaluation that showed Sonora to be the leader in educational achievement”.

“In this regard, I believe that one successful endeavor that should be publicized and replicated is the New School Program that was designed based on the setting up of the State Evaluation System in 2004 in compliance with our state’s law governing evaluation and as result of the establishment of the INEE. We need to preserve and strengthen the essential features of the said support program for school managers in order for Sonora to once more become a leader in educational quality”. €



You can find out more about the Ministry of Education and Culture of the State of Sonora at: <http://www.sec-sonora.gob.mx/portal/index.php>



## Effective use for improvement; what can we do to make it happen?

“For the promise inherent in evaluation as an instrument for improvement to be effectively realized, it not only needs to be created in a thorough, fair and timely manner, but also must actually be used by all those involved for the clear, explicit purpose as serving as evidence to orient and inform decisions about improvement in each area of practice”, says the author of the following article, who proposes a tool for analyzing users within a policy aimed at foster effective use.

TERESA BRACHO GONZÁLEZ  
Member of the Board of the INEE  
[teresa.bracho@inee.edu.mx](mailto:teresa.bracho@inee.edu.mx)

By effective uses of evaluation, one means deliberate actions implemented by the different recipients or users of information (i.e. protagonists), who -to different degrees and for different purposes- use the results, methods or ideas derived from educational evaluation as relevant input on which to base decisions and actions aimed at bringing about improvement in their particular spheres of action.

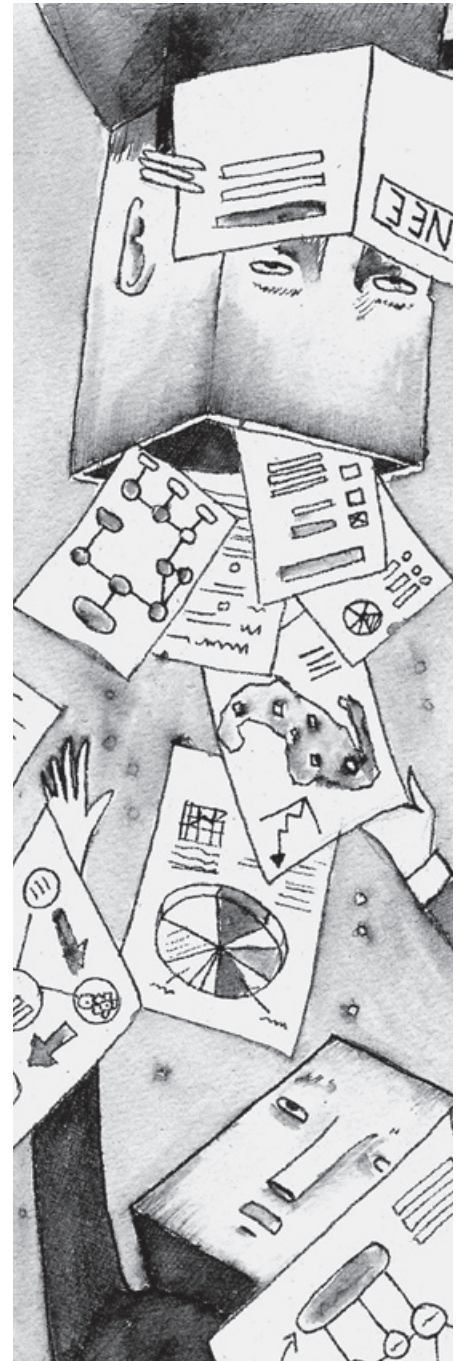
In its early stages, in the late 1960's, it was considered that the sole function of evaluation, and of the information stemming from it, was to impact the particular program evaluated. However, as Carol Weiss explains, we are now aware that evaluation has other dimensions, and that many other aspects of it besides its results -e.g. its ideas and visions, theoretical frameworks, methodological procedures and even its design- can also be used, since “use encompasses a

broad array of effects by multiple classes of users” (Weiss, 1998).

In practice, we have gradually progressed from creating a small range of evaluation reports that were of limited relevance or use to potential users, and merely concentrated on complying with standards or satisfying requirements, to producing evaluations that increasingly take the said users into account, notwithstanding which we are still very far from achieving optimal evaluations that take stock of all the relevant factors and take stock of all the possible users.

Hence, we still have a long way to go when it comes to fostering meaningful uses of evaluation aimed at improving the education system. Often, as specialists in evaluation, we have mainly worried about the scientific quality of our work, and this has led us to focus on things such as the rigor of our measuring tools and their analytical methods, and the problems inherent in ensuring data integrity and security. Of course, these things are our responsibility and constitute prerequisites for achieving improvement, but they do not suffice to bring it about. We must also view of policy as something authentically public and seek to involve all the protagonists, at every level, in order to solve problems. There are implicit beliefs or suppositions about the use of information stemming from educational evaluation, which, if we examine them from viewpoints such as that of public policymaking, could tell us some interesting, relevant things about the successes and shortfalls of our work as evaluators of the education system so far.

For our evaluation to truly become a factor in the improvement of the said system, it invariably needs to be used at all levels and by all the protagonists in their particular spheres of action. But how do we ensure that this happens? Do we need to do something different in order to ensure that evaluation is really useful? In my opinion, we evaluators need to change our viewpoints and practices in at least the following three ways: we need to (a) take stock of the uses of evaluation starting at the planning stage (what I call *stage zero*), (b) transcend the traditional outlook that underlies our behavior as evaluators, and (c) address the subject of the uses of evaluation aimed at “intervening in public policy”, by:



### a) Considering the topic of the uses of evaluation starting from the latter's conception – i.e. at Stage Zero

This topic must be put on the agenda – i.e. it must be one of the main concerns of evaluators from the moment when they start to design an evaluation, which means changing the two typical practices. First we must cease thinking about the uses of evaluations

after they have been completed, only then -and not before- deciding which types of reports, publications and methods of dissemination to use, since this way of thinking leads us to lose an opportunity that cannot always be recovered later, given that it may not be possible to adapt the data gathered during the evaluation process to suit uses that were not envisaged at the design stage. Second, we must stop believing that our job has ended when the results are published and disseminated, assuming that the uses needed to change the system at its different levels will arise spontaneously as a result of interest, understanding and assimilation on the part of the various protagonists who are interested in the workings of the education system. In short, I propose that communication about evaluation extend beyond the publication and dissemination that have traditionally been carried out in Mexico.

#### **b) Changing the “traditional” outlook that underlies our behavior as evaluators of the National Education System (Spanish acronym: SEN)**

Linked to point *a)* above, one can assert that, precisely due to the specialized nature of our task, we have fallen prey to a set of fallacies that we have not explicitly addressed in the course of our work, and hence not talked about enough. If we want to foster the actual use of evaluation for purposes of improvement, we must, at the very least, question the implicit assumptions that:

1. the important thing is to produce high-quality evaluations, with our main challenge as evaluators being to create information of the required quality, which can be achieved by developing highly rigorous and scientific instruments.
2. publication and dissemination ensure access and divulgation, since the mere publication and dissemination of scientifically controlled evaluation results makes them accessible and understandable to those involved in the SEN.
3. users are fully able to understand the information that is produced, since educational evaluations are mainly designed to be used by specialists.
4. due to his/her training and experience, the evaluator is the one who knows best what should be evaluated, what can be evaluated, and how to do so.

Some of the aforesaid assumptions may have worked when evaluation was seen as being limited to the function of accountability, providing information to top officials about the overall progress of the education systems they administered. They may have been reasonable when respect for -and trust in- external evaluation, and the incursion of professional evaluation specialists into public affairs, had to be engendered. However, now that we have changed, and enriched, our way of looking at education systems, and in a context where access to information has become a right per se, any evaluation practice that mainly rests on the aforesaid, or similar, assumptions prevents us from achieving the longed-for link between evaluation and educational improvement, and should be eschewed.

#### **c) Seeing the use of evaluation as an intervention in government policy**

Once it has been accepted that use should be taken into account starting at the planning stage of evaluation, and we have adjusted our outlook to match our practice, the other important point is that we need to conceive of the use of evaluation as a form of intervention in government policy, rather than a matter of haphazard, spontaneous response to the interests of the various users of the evaluation data. Though those of us who form the National Educational Evaluation System (Spanish acronym: SNEE) need to have a broader discussion about the matter, I propose at least two topics to be covered during the said debate.

First, if we see the matter of use as a policy in itself, rather than a spontaneous, fortuitous consequence of evaluation, as already mentioned, *then the non-use or rejection of evaluation findings, or the failure to understand them, constitute the specific problem that the said policy should solve*, and second, as is the case with all suggested policies, in order to be deemed to constitute public ones, those we are talking about here require rigorous, participatory analysis of their potential usefulness and of the needs and expectations of all the members of the SEN *who are able to play a part in the latter's improvement*, to which end I go on to propose a simple tool -one of many possible ones- that can be used to structure the aforesaid processes of needs identification and categorization, dialogue and building of meaning related to evaluations.

#### **Tool for analyzing users as part of a policy aimed at fostering effective use.**

As already affirmed, it is essential that potential users be considered, and all the more so in the case of educational evaluation. If we wish to inform policymaking in its different areas, and at its different levels, we must produce evidence expressed in the languages and forms that are appropriate for the users at different levels, taking into account the said people's actual ability to take on board, interpret and use the evidence in question within their respective spheres of action.

Above all, it is essential that those who produce the evidence -i.e. the evaluators- be aware that the aforesaid consideration of the characteristics and needs of the target users must be the starting point for designing strategies capable of producing results that can foster uses that lead to improvements in our education system.

As a first step, I propose using the “observation areas” (Fig. 1) customarily used by researchers to measure education-service quality, which are clearly summarized by the Latin-American Laboratory for the Assessment of Educational Quality (Spanish acronym: LLECE), an initiative coordinated by UNESCO, and used by the latter for its achievement tests (i.e. the First Regional Comparative and Explanatory Study (Spanish acronym: PERCE), the Second Regional Comparative and Explanatory Study (Spanish acronym: SERCE), and the Third Regional Comparative and Explanatory Study (Spanish acronym: PERCE)).

Then, as a second step, I propose a “mirror” analysis (Fig. 2) taking the same agents and protagonists as in Fig.1 -i.e. students, teachers, school principals and authorities- as the potential initial users of any evaluation of educational quality (not forgetting that this second analysis is based on an evaluation of achievement, since, if the aforesaid people are the ones who play a part in determining quality, being able to take steps to improve classroom education, they are also the ones to whom we should primarily address our results and keep foremost in our minds when carrying out any evaluation.

Though simple, this concept can enrich the first approach at defining the products that spring from evaluation results. However, since they should be defined in the context of the National Educational Evaluation System (Spanish acronym: SNEE), I have

inserted question marks in the four sections of Fig. 2. As a matter of priority, we need to define the type of evaluations to be used and determine how the information yielded by them can be utilized to foster high educational quality via consultations with the main potential users, among other things.

It should be stressed that this matter is of the utmost importance strategically, especially in the context of the reforms currently underway in the SEN, since, in order to exploit the potential of *evaluation as a tool for improvement*, the said evaluation must not only be produced in a rigorous, fair and timely manner, but also actually used by all the players involved -each at his/her/its respective level- for the clear, explicit purpose of providing guidance so that the said players can make decisions aimed at achieving improvement in their respective spheres of action.

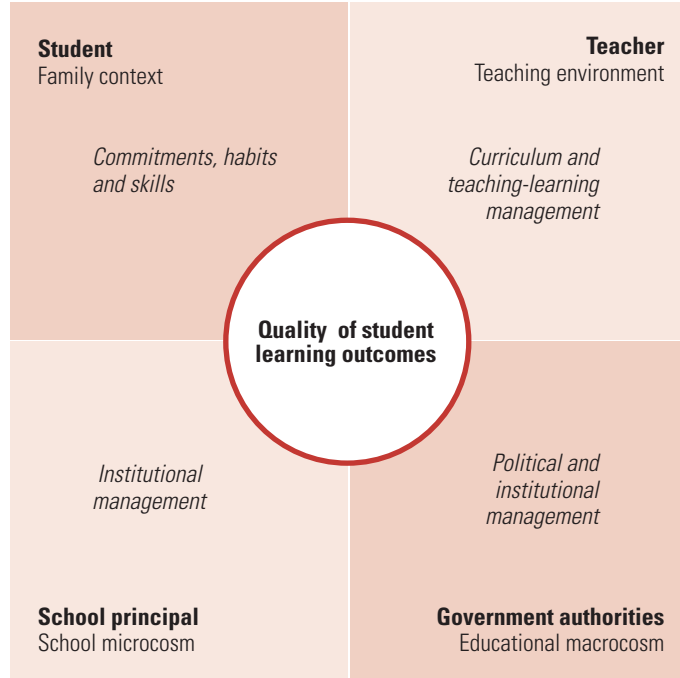
The use of educational evaluation -a topic that differs from, but is closely related to, that of the creation or production of evaluation- is a subject that merits research in its own right, since the effective uses that are desired -i.e. ones that lead to improvements in the education system- will not occur spontaneously or automatically to the different existing or potential recipients of evaluation findings. We on the supply side have some very important leeway in the endeavor to foster greater, more thoughtful demand and thus transform desired uses into improvement strategies. It is time for us to embark on this task. €

**References**

Latin American Laboratory for the Assessment of the Quality of Education, LLECE. (2000). *First international comparative study of language, mathematics, and associated factors for students in the third and fourth grade of primary school*. Second report. Santiago de Chile: UNESCO. Available at: <http://unesdoc.unesco.org/images/0012/001231/123143s.pdf>

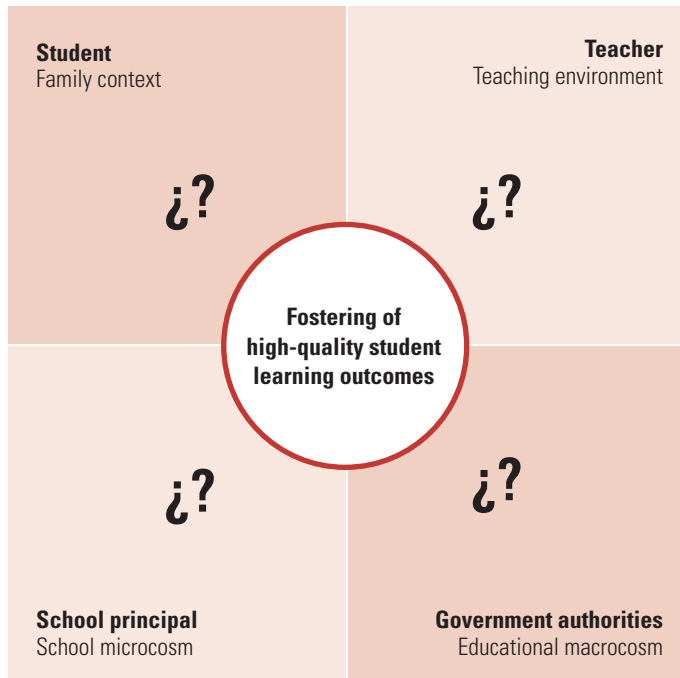
Weiss, C. H. (1998). Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation*, No. 1, pp. 21-33.

**Fig. 1: Observation areas used in educational research**



Source: Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Ilece), 2000:12.

**Fig. 2: Tool for identifying the potential users of evaluations in an effective-use policy; protagonists in their contexts**



Source: Author-producer based on LLECE 2000:12.



## NAUTICAL LETTER

## A NATIONAL EDUCATIONAL EVALUATION POLICY FOR PROMOTING IMPROVEMENT

**Topic:** A National Educational Evaluation System (Spanish acronym: SNEE) that helps to improve educational quality in Mexico

**Main aim:** To give a brief referential overview of the creation and current status of the SNEE, and the outlook for the latter, that will be of use to policy-makers and serve to plot out the path to be followed with regard to educational evaluation, based on the Governing Document of the National Educational-Evaluation System (SNEE-GD).

### 1. The SNEE-gd as a starting point

Arising from reflection, the exchange of ideas and the reaching of consensus, the SNEE-GD provides a long-term panorama that both sets out to make good on the right to education and constitutes a new agreement between the INEE and the federal and local education authorities.

Since 2014, we have been holding the Dialogues with the Educational Authorities, intended to be a forum for direct contact and high-level discussion between the INEE's Board of Governors and the education authorities about the implementation of the Educational Reform and the chal-

lenges inherent in the evaluation of the National Education System (Spanish acronym: SEN), and hence to explicitly help to build a shared, interlinked PNEE agenda.

The efforts made by the INEE's in 2015 to carry forward the building of the PNEE focused on (i) analysis and open discussion with those involved in the SNEE about the appropriateness of making a right-to-high-quality-education emphasis the PNEE's guiding principle; (ii) the definition of the mainstays of the PNEE; (iii) the stipulation of the purviews and responsibilities of each of the authorities that make up the SNEE based on a legal analysis of the powers and competencies of the different entities involved in educational evaluation; and (iv) discussion, in the Conference, about the SNEE's Governing Document with the federal and state-level education authorities.

### 2. The advantages of having a National Educational Evaluation Policy

The PNEE is envisaged as a tool for bringing about educational improvement that will apportion the different evaluation tasks among the members of the SNEE and guide their efforts, clarify which legal powers have been conferred on each of the said

members and make it possible to coordinate efforts, at both the national and local levels, to establish the aims of evaluation and monitor its short- and medium-term progress<sup>1</sup>.

In this regard, the PNEE constitutes an effort to give educational evaluation a bigger role as an interface between the said evaluation and the use of its results to improve education.

It should be stressed that, when referring to the SNEE, we mean the three main components that it will embody – i.e. (i) the SNEE's Governing Document<sup>2</sup>; (ii) the State-level Evaluation and Educational-Improvement Programs (Spanish acronym: PEEME)<sup>3</sup>, and (iii) the SNEE Medium-Term Program<sup>4</sup>.

### 3. How can the PNEE be used to improve education?

The PNEE will both ensure that the evaluations carried out throughout Mexico, at both the federal and local levels, are objective, transparent, accurate and unbiased, and also make it possible to create institutional channels for coordination and cooperation that make the government actions undertaken in the context of the SNEE more effective, leading to better liaison between

the Areas Responsible for Evaluation (Spanish acronym: ARE's) and the federal and state-level authorities, and also between the said ARE's and other ARE's, Higher Education Institutions, international organizations and NGO's, thus furthering the carrying out of their projects and activities.

It will also produce institutional and intervention mechanisms for improving state-level education services via the use of evaluation results, making it possible to come up with improvement strategies mainly based on the said results.

Finally, through the guidelines, the PNEE will provide guidance which, while not breaking the law, will make it easier to comply with the legal mandates, informing educational policy-making in order to raise the quality of education. It should be stressed that the said guidance can be given at the request of the education authorities.

#### 4. What will the PNEE's next step be in 2016?

##### State-level Evaluation and Educational-improvement Programs

The PEEME's mentioned above will be designed by the states with the INEE's help, and this task requires collective work on the part of the state-level education authorities responsible for both elementary and lower-secondary education in order to put together a single state-level program that covers both levels.

Diagram 1. Components of the PNEE



#### National Educational-Evaluation Policy

##### Medium-term SNEE Program

Stemming from a review, analysis and integration process, the Medium-term SNEE Program (SNEE-MTP) will put together the basic components of the SNEE developed by each state, which will indicate which actions are to be taken at the national and state levels respectively, as well as stipulating the activities, deadlines, people in charge, and goals pertaining to the federal and local education authorities and the INEE.

You can find the full version of the Document Governing the National Educational-Evaluation at: <http://goo.gl/YMjkKP>

- 1 Projected to 2020.
- 2 Includes the central evaluation-policy tenets that orient and organize the work path, lines of action and perspective for 2020 and address the constitutional obligation to raise the

quality of compulsory education and make it more equitable via cooperation among the protagonists and institutions involved in the SNEE.

- 3 This planning tool for analyzing and orienting educational evaluation and improvement in the states consists of an institutional program that will facilitate the setting up of projects, aims and actions pertaining to evaluation and educational improvement and goals, organizing one or more evaluation projects that the state deems pertinent based on its diagnostic study.
- 4 Will set out (i) to define the legal, conceptual and methodological foundations and also the implementation paths for carrying out the evaluation processes defined in the SNEE, and (ii) to establish the objectives, lines of action and calendar for such evaluation processes as are eventually defined.

## 2) Roadmap: stages between January and July, 2016



Note: The INEE will provide feedback at each stage and the stage education authorities will make the pertinent adjustments.

#### Would you like to learn more about the PNEE?

Look for the special supplement of this *Gazette*: "The SNEE and the National Educational-Evaluation Policy: Challenges and Perspectives".

# ROADMAP FOR DEVELOPING THE 2016 PEEME'S

## 1) The regions



## 3) Calendar

Activities and deliverables	Date
First version of the Diagnostic Study	January
Feedback on the Diagnostic Study from the INEE	February
Adjustments to the Diagnostic Study	February
<b>REGIONAL MEETING: FIRST STAGE OF PEEME</b>	February
First version of the List of Problems	March
Feedback on the List of Problems from the INEE	March
Adjustments to the List of Problems	March
First version of Projects and Aims	April
Feedback on Projects and Aims from the INEE	April
Adjustments to the Projects and Aims	April
<b>REGIONAL MEETING: SECOND STAGE OF PEEME</b>	April
Second meeting of the INEE-SEP PEEME Development Committee	May

Activities and deliverables	Date
First version of planned Actions	May
Feedback on planned Actions from the INEE	May
Adjustments to planned Actions	May
First version of Aims	June
Feedback on Aims from the INEE	June
Adjustments to Aims	June
<b>REGIONAL MEETING: THIRD STAGE OF THE PEEME</b>	June
First version of Schedules and Benchmarks	July
Feedback on Schedules and Benchmarks from the INEE	July
Adjustments to Schedules and Benchmarks	July
Third Meeting of the INEE-SEP PEEME Development Committee	July
Fourth Meeting of the INEE-SEP PEEME Development Committee	October

## The meaning and use, for the improvement of Education, of the results of the Evaluation of Basic Teaching-Learning Conditions\* and the Evaluation of Learning Outcomes in Elementary Education\*\*

JORGE ANTONIO HERNÁNDEZ URALDE  
Head of the National Education System's Evaluation Unit (Spanish acronym: UESEN) of the INEE  
[johernandez@inee.edu.mx](mailto:johernandez@inee.edu.mx)

ANDRÉS EDUARDO SÁNCHEZ MOGUEL  
General Director for the Evaluation of Educational Results of the National Education System's Evaluation Unit of the INEE  
[asanchez@inee.edu.mx](mailto:asanchez@inee.edu.mx)

RAQUEL AHUJA SÁNCHEZ  
General Prospectus Director for the National Education System's Evaluation Unit of the INEE  
[rahuja@inee.edu.mx](mailto:rahuja@inee.edu.mx)

### Introduction

In late 2015, the National Institute for the Evaluation of Education (Spanish acronym: INEE) published the preliminary results of the Evaluation of Basic Teaching-Learning Conditions (Spanish acronym: ECEA) and the Evaluation of Learning Outcomes in Elementary Education that forms part of the National Plan for the Evaluation of Learning Outcomes (Spanish acronym: Planea). Although the said evaluations have different, independent, purposes, since ECEA sets out to ascertain the extent to which Mexico's compulsory-education institutions have the basic wherewithal to function, while Planea

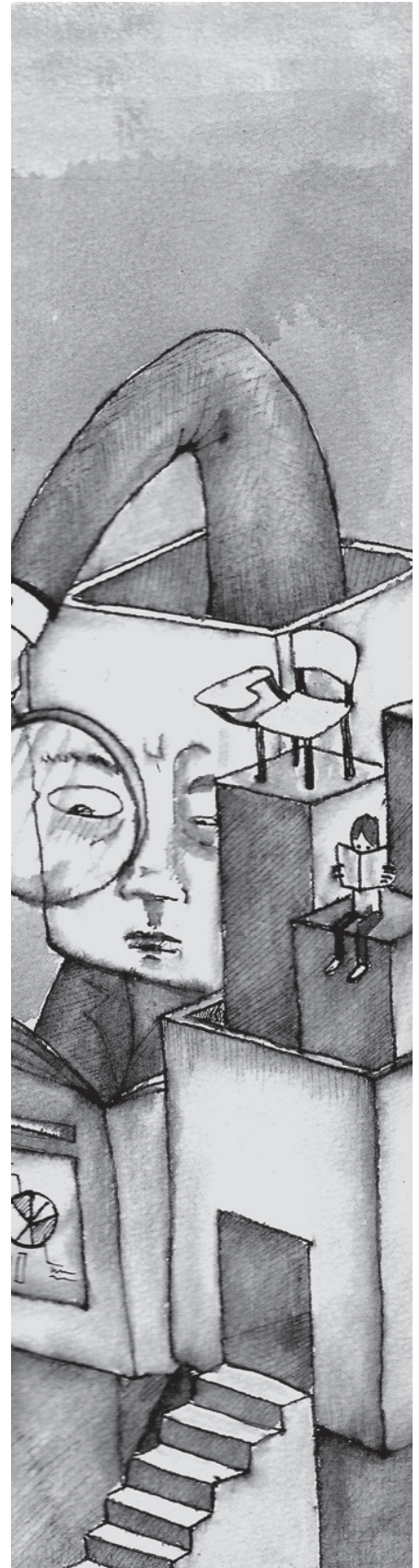
endeavors to measure students' mastery of a set of essential learnings at different stages of the compulsory-education process, the main results of both of them, which are outlined in this article, show aspects of our country's educational *status quo* that are closely interlinked with each other. After first presenting the main Planea results regarding learning outcomes in the sixth year of primary school, we will go on to outline the basic conditions in Mexican primary schools shown by the ECEA results, and, finally, we will venture some observations about the overlaps between the said two studies and the uses to which the information yielded by them can be put.

### Learning outcomes at the primary level

Planea was administered by the INEE, at the end of the 2014-2015 school year, to a representative sample of sixth-year primary-school pupils that was used as a basis for extrapolating the countrywide results that we will comment on below. The Ministry of Public Education (Spanish acronym: SEP) administered it in all the other primary schools in Mexico so as to furnish each school with a report on the results achieved by its pupils, duly contextualized. The tests that make up the said exam cover developmental areas on which many other types of learning are based, such as Language and Communication and Mathematical Reasoning, and which hence constitute two areas of the curriculum that serve as yardsticks or proxies for other areas, it being difficult, for example, to imagine students performing well in the sciences if their linguistic and mathematical skills are weak.

Generally speaking, the Planea results at the primary level correlate with the low performance levels in the National Education System that previous evaluations such as the Examination of Educational Quality and Achievement (Spanish acronym: EXCALE) and the National Examination of Academic Achievement in Schools (Spanish acronym: ENLACE) had already been indicating.

Almost half (49.5%) of the students placed in Achievement Level 1 in Language and Communication, meaning that they face serious limitations when endeavoring to carry out the basic language tasks such as understanding expository<sup>1</sup> and literary texts



\* Spanish acronym: ECEA

\*\* Spanish acronym: Planea

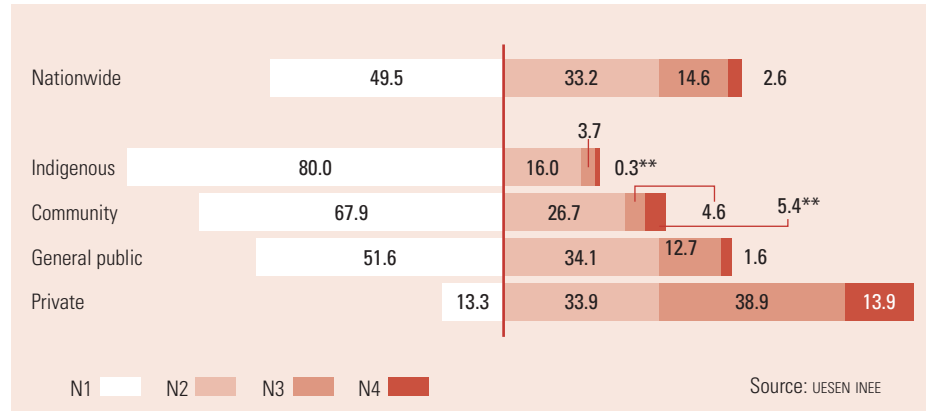
that they are required to master as part of the primary-school curriculum, and comparing the communicative aims of different types of text, with only a third (33.2%) of the students who placed in Level 2 -pertaining to an indispensable minimal mastery of key curricular learnings- being able to carry out the said tasks.

Around 15% of sixth-year primary-school pupils placed in Achievement Level 3, meaning that they can carry out inferential tasks such as understanding a metaphor in a story, writing topic sentences that express the main idea of a text and summarize the author's intentions, and identifying data, arguments and opinions, while only 3% of the said pupils are in achievement level 4, which means that they can deploy top-level cognitive skills to understand argumentative texts, deduce how an interview is structured, etc..

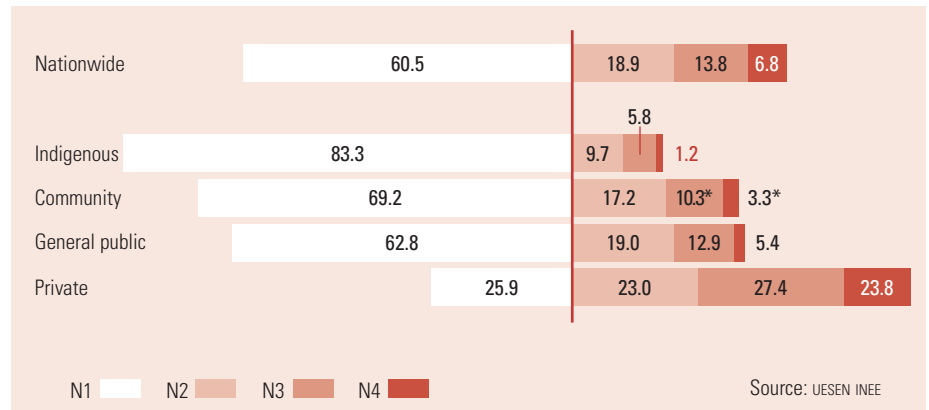
Just over 60% of students placed in Level 1 for Mathematical Reasoning, meaning that they still have not acquired the fundamental abilities pertaining to the said, such as the capacity to carry out basic operations (i.e. addition, subtraction, multiplication and division) using natural numbers, identify geometrical features such as height, parallelism and right angles in simple shapes, or solve problems involving perimeters, while 19% of all sixth-year pupils placed in Level 2, meaning that can indeed perform the aforesaid tasks. Around 14% placed in Level 3 and are also able to solve mathematical problems involving decimals, compare fractions and multiply them by natural numbers, and identify geometric progressions using rules, etc., while only 7% of the students who complete the sixth year of primary school managed to place in Level 4, meaning that, as well as having the aforesaid competencies, they can solve addition problems using fractions, divide and multiply fractional numbers by natural numbers, calculate areas and work out averages and means.

In short, a very significant portion of the students who complete primary school in our country is not achieving satisfactory learning outcomes. By splitting the student population's learning outcomes into four levels, we can perceive the most significant differences among students who attend different types of school, as shown in Graphs 1 and 2 below, which depict the said differences in the respective fields of Language and Communication and Mathematical Reasoning. As can be seen in these graphs,

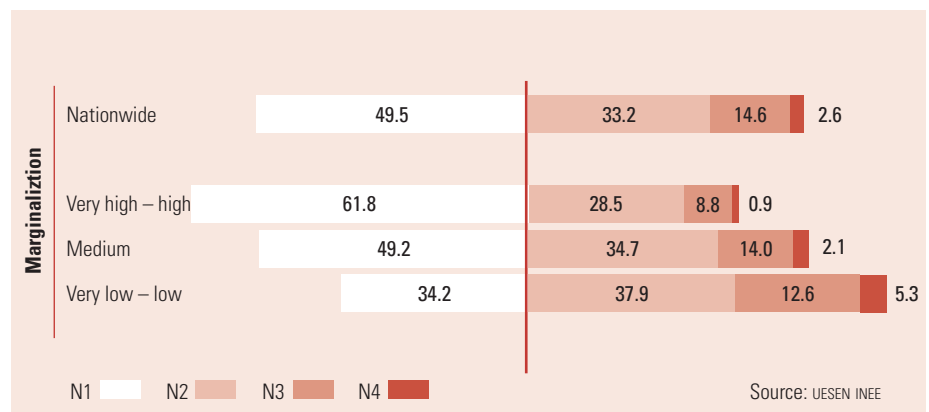
**Graph 1. Percentage of students per achievement level in Language and Communication, nationwide and for each type of school: sixth grade of primary school, 2015**



**Graph 2. Percentage of students per achievement level in Mathematical Reasoning, nationwide and for each type of school. Sixth grade of primary school, 2015**



**Graph 3. Percentage of students per achievement level in Language and Communication according to the level of marginalization. Sixth grade of primary school, 2015**



the types of school located in the most vulnerable areas -i.e. indigenous and community schools- have much lower achievement levels than other public and private schools, with 8 out of every 10 pupils in indigenous schools and 7 out of every 10 pupils in community schools placing in achievement level 1 in both Language and Communication and Mathematical Reasoning.

This tendency for the most vulnerable populations to achieve the worst learning outcomes also becomes evident when one analyzes the distribution of achievement levels taking stock of the levels of marginalization of the schools' locations. Graph 3 below, which shows the distribution of achievement levels in Language and Communication according to local marginalization levels, shows that those locations with high or very high levels of marginalization have the lowest achievement levels, while the highest achievement levels are found in those locations with low or very low levels of marginalization, with a similar trend being found for Mathematical Reasoning.<sup>2</sup>

All the aforesaid data indicate that the inequality in learning outcomes is largely due to the way the education system works, offering different types of service to different populations and, as far as can be seen, helping to aggravate social inequality rather than diminishing it. Some of the results showing this other side of the coin are presented below.

### Basic conditions in primary schools

ECEA, an evaluation of operating conditions in schools that is carried out every four years, measures the extent to which the said institutions have the basic (not necessarily optimal) resources and processes for their pupils to learn, the said basic wherewithal being those that every school, and all Mexican students, should have – i.e. physical resources such as infrastructure, furniture and teaching materials, teachers and management staff who work collaboratively, parent participation, effective use of time, and conviviality among pupils.<sup>3</sup>

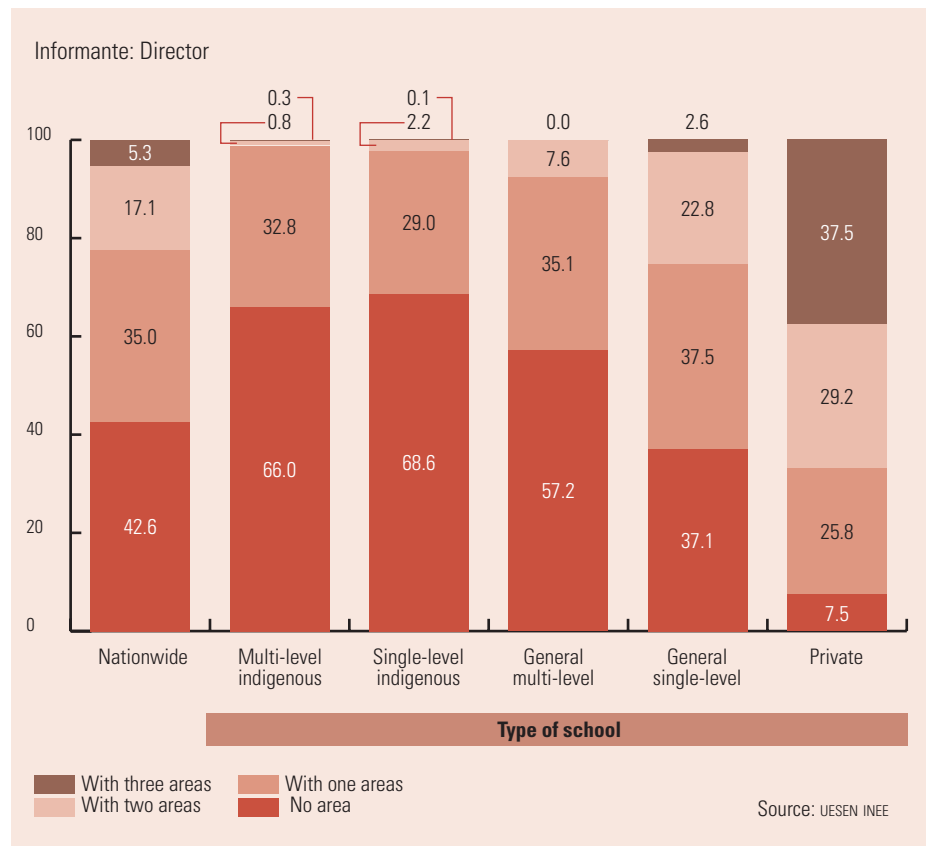
ECEA was first administered in 2014 to a representative sample of our country's primary schools that comprised six types of institutions – i.e. general single-level public primary schools, general multi-level public primary schools, single-level indigenous schools, multi-level indigenous schools, community schools and private schools.<sup>4</sup> The principals of each school, plus teachers

and students from the 4<sup>th</sup>, 5<sup>th</sup> and 6<sup>th</sup> grades and a representative from the board of the Parents' Association, acted as informants. Below, we present some of the results that indicate differences in the educational conditions of Mexican students, and show that the biggest shortfalls were in schools located in vulnerable contexts.<sup>5</sup>

To support learning and the coverage of the curriculum, the schools have libraries, media rooms and multiple-use rooms as well as classrooms, with 60% of all Mexican

schools possessing at least one of the aforesaid facilities, but only 5% having all of them (See Graph 4). Almost none of our country's public schools has all three of the aforesaid educational-support areas, all three of which are mostly found in the general single-level schools, of which 2.6% possess them, as compared with 35% of all the private schools. Two out of every three indigenous schools have none of these three areas, while more than half (57.2%) of the general multi-level schools, and over a third (37.1%)

**Graph 4. Percentages of schools with 1, 2 and 3 educational-support areas respectively: nationwide and by type of school<sup>6</sup>**



of the general multi-level schools lack them.

While 81% of all Mexican state schools have the student-only bathrooms that are essential for security and privacy (Graph 5), this figure decreases in line with the degree of marginalization, dropping to just 44.4% in the community schools.

Classroom furniture is a basic part of school equipment. Nationwide, all the children in 78.2% of all the groups have a desk

and seat where they can sit and write in good or fair condition, but again this is more the case in the private schools (99.5%) and the community schools (82%), where the National Council for the Fostering of Education (Spanish acronym: CONAFE) is directly responsible for school equipment. Next come the general schools (68.3% to 78.3%) and finally the indigenous schools (53.5% to 59.2%) [Graph 6].

It can be seen that basic educational materials such as free textbooks (FT's) are not opportunely distributed in sufficient quantities in all the different types of school. Though 80% of the single-level general schools and the private schools received FT's before the start of the school year, the figure for general multilevel and indigenous schools is around 70%, and just over half (51.7%) for community schools, it being important to stress that, almost three months into the 2014-2015 school year, a significant number of community schools (9.4%) reported that had still not received any FT's. Moreover, three months after starting classes, not all the children had a complete set of FT's, as can be seen in the following table.

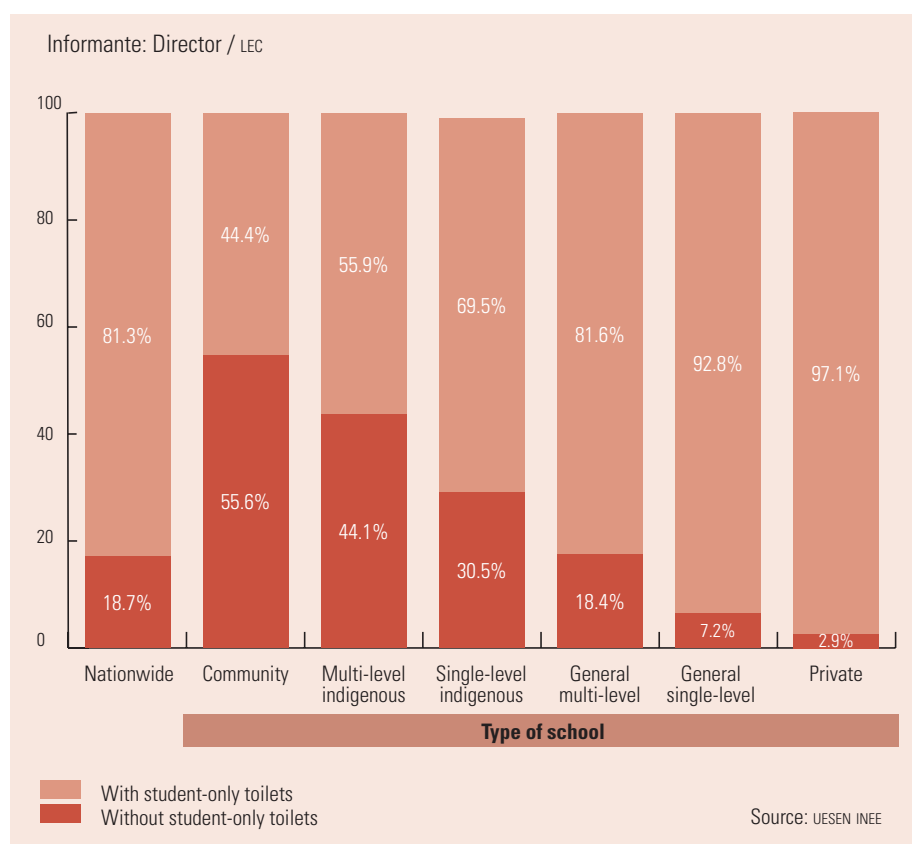
The results presented here are just an example of the inequality in children's study conditions. The education system needs, at the very least, to ensure that there is a general *shared base level* for all schools and hence all students, though, from an equality viewpoint, the schools and children in the poorer locations should receive not just the same, but more, in order to be able to achieve their learning targets.

### Some final thoughts

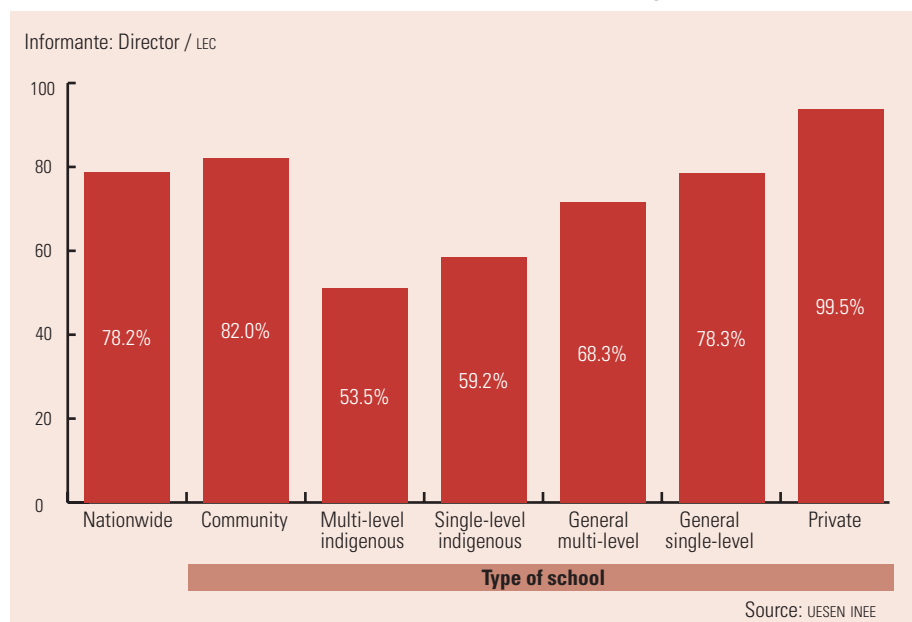
Learning-outcome levels are also lower in the types of school that function in the most precarious conditions and hence offer less wellbeing and learning opportunities to their students, who come from the poorest environments, as reflected in their results and in the inequalities among the courses offered. The schools in these conditions face a bigger challenge than the, both because they cater to impoverished populations, and also because they offer less favorable study conditions.

Both ECEA and Planea clearly show that we urgently need to improve the educational services provided to the most vulnerable populations, as a matter of priority, aiming to ensure not only that all our children attend school, but also that they all learn. In this regard, it is germane to ask the authorities responsible for managing primary education, to respond, with the participation of state-level ministries of education, groups of researchers, teachers, parents and other people involved in education, to the following questions: (i) How can we ensure that all students achieve the learning outcomes that are indispensable at the primary level? (ii) Should the curriculum be structured

**Graph 5. Percentages of schools with working toilets exclusively for students: nationwide and by type of school**



**Graph 6. Percentages of schools whose students have a desk and seat where they can sit and write: nationwide and by type of school**



**Table 1: Percentages of schools that received Free Textbooks before the start of the 2014-2015 school year or at the start thereof and percentages of groups in which the students had a full set of FT's in November, 2014.**

Type of school	% of schools that received Free Textbooks before the start of the 2014-2015 school year or at the start thereof	% of groups in which the students had a full set of FT's (November, 2014).
Community	51.7%	30.3%
Multi-level indigenous	70.9%	48.3%
Single-level indigenous	67.0%	43.1%
General multi-level	72.6%	54.5%
General single-level	80.6%	64.9%
Private	80.0%	77.7%

Fuente: UESEN INEE

around essential common primary-level learning aims and others that adjust to different contexts? (iii) How can we strengthen initial and in-service training for those who teach more needy populations? (iv) How can we ensure that all the schools belonging to the National Education System comply with basic operating norms and offer the same learning opportunities to all? (v) How do we establish special incentives for teachers facing higher challenge levels? and (vi) How do we reconcile the need to use resources efficiently with the fact that the universal right to high-quality education implies affirmative action that requires us to spend more on the most vulnerable populations?

We still need to ascertain the extent to which the ECEA and Planea results are known, and used, by school communities. The implementation of Planea 2015 included the gathering of information at both the national and state levels that was published by the INEE, and also of information about each school, which was published by the Ministry of Education so that the School Technical Committees and the School Social-participation Committees might analyze it and reflect on local strengths and opportunities. Undoubtedly, we need to make several adaptations and improvements in the dissemination of the results at the pri-

mary level in order to promote their use for purposes of improving our schools.

For its part, ECEA uses a *basic school-operation framework* that includes 68 conditions split into 7 fields and 21 dimensions, as well as 160 benchmarks that can be analyzed by school communities and education authorities at the school-zone, sector-head, and state-education-authority levels, etc. for the purpose of evaluating school conditions and planning actions to improve them.<sup>8</sup>

In this article, I have tried to how the evaluation results can provide information that is pertinent to different protagonists and inform decision-making in different spheres, making different requirements of each of them. Evaluation is useful to the extent that the responsible authorities ponder its results and use them to make decisions.

Those wishing to find out more about the evaluations mentioned in this article should visit the following microsities via the INEE portal. €

- 1 Texts whose main aim is to transmit information and disseminate knowledge.
- 2 It is not possible to include all the information that is available about the sixth-year-primary Planea results for 2015 in this article. The complete results can be consulted at: <http://www.inee.edu.mx/index.php/planea>

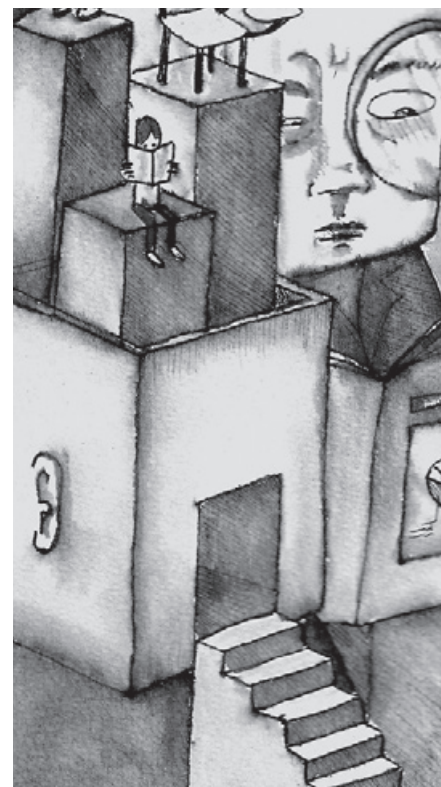
- 3 More information about evaluation can be found at: <http://www.inee.edu.mx/index.php/proyectos/ecea>
- 4 It should be stressed that most of the multi-level general schools -i.e. schools where a single teacher handles groups whose students are at different levels- and the indigenous and community schools are situated in rural locations with high or very high levels of marginalization, while the general single-level schools and the private schools are located in urban areas with medium, low or very low marginalization levels.
- 5 The results were obtained via self-administered questionnaires plus a personal interview with the Parents' Association representative.
- 6 Educational-support areas, including libraries, multiple-use rooms and media rooms.
- 7 Community schools are not included in the graph.
- 8 The basic operating framework for primary schools that was used in ECEA 2014 can be consulted at: <http://www.inee.edu.mx/index.php/proyectos/ecea>



ECEA: <http://www.inee.edu.mx/index.php/proyectos/ecea>



Planea: <http://www.inee.edu.mx/index.php/planea>





## ROADMAP

## The INEE guidelines: building bridges between evaluation, uses and educational improvement

The INEE issued its first guidelines in 2015, after being empowered to do so by the 2012 Educational Reform, which links evaluation to improvement. The author of the following article gives a brief summary of the road that the aforesaid Institute mapped out for the creation of the said guidelines, stressing both the key role played by policy and program evaluations aimed at remedying the problem, shortfall or weakness that is being addressed, and also emphasizing the need to hear the viewpoints of different protagonists.

ARCELIA MARTÍNEZ BORDÓN  
General Director of Guidelines for Educational Improvement  
Regulation and Educational-policy  
Unit of the INEE  
[amartinezb@inee.edu.mx](mailto:amartinezb@inee.edu.mx)

All government policies are responses by the State to social problems or to situations that are deemed unsatisfactory or unacceptable (Subirats, Knoepfel, Larrue and Varone, 2008). Drawn up in many ways, but resulting, above all, from pressure applied by different protagonists for the existence of a social problem to be acknowledged, policy agendas may also receive input from evaluations and research, and, indeed, the INEE's main aim is that educational evaluation help inform decision-making about how to raise the quality of education.

The Educational Reform proclaimed in December of 2012 conferred autonomy on the said Institute and empowered it to issue policy recommendations -i.e. "guidelines"- aimed at informing decision-making about educational policy and thus serving as a bridge between evaluation and educational improvement.

In this article, which contains some reflections about how to create the aforesaid INEE guidelines and about their potential for influencing decisions aimed at improvement, we first summarize part of the debate regarding the things that prevent evaluation from being used and transformed into government policy, and then go on to mention two key aspects of the drawing up of guidelines that seek, precisely, to achieve such transformation – namely, the participatory defining of the problem being addressed and the possible solutions to it, and the weighing of the scope and shortfalls of government action at the state level.

### Difficulties inherent in the research-evaluation pairing and its use

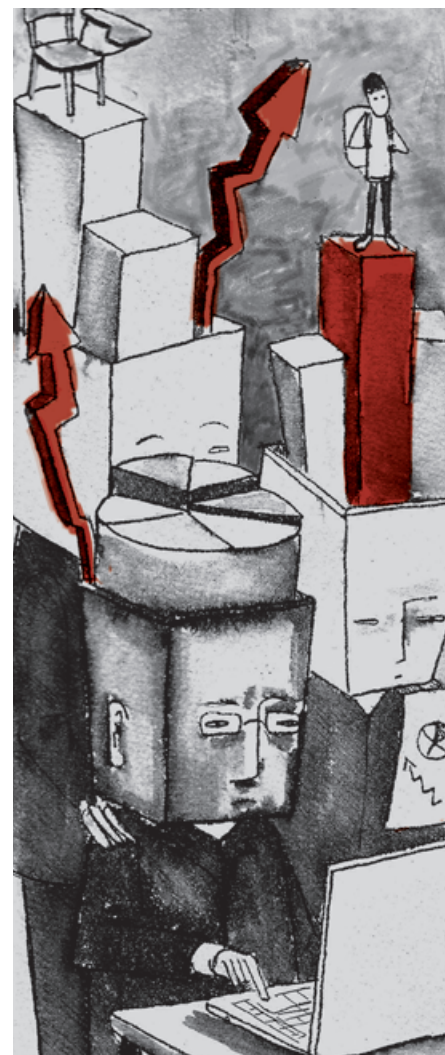
Carol Weiss holds that, though the linear model whereby pure and applied research are turned into policy presupposes the existence of consensus between researchers and policymakers about the aims being pursued, nothing could be further from the truth, since research and policymaking run parallel to each other, occasionally coinciding, but only tangentially, and, while academic debate helps to establish the paradigms for the discourse frameworks that govern political action, it rarely defines the public-policy agenda.

While it is assumed that, if research and evaluation comply with certain standards, the rest is history, in fact, though playing a key role in informing decision-making, the high quality of the said research and evaluation is no guarantee that they will be taken stock of by some user groups, since other things, such as political negotiation and consensus seeking and forming, play a bigger role in the use of evaluations (Moreles, 2010).

Another important factor preventing research and evaluation from being used is the fact that, in most cases, the said research or evaluation, and publication of the findings

thereof, take too long to serve political purposes, so that decision makers, faced with the need to act and respond rapidly, use whatever available information helps them to underprop their decisions.

Other factors that determine whether research and/or evaluation will be used are the profile of the researchers/evaluators, the prestige of the institution to which the latter belong, and the type of research and/or evaluation, added to which is the fact that quantitative studies still enjoy more prestige than qualitative or meta-analytical ones, and that overview studies still enjoy more prestige than studies on a single topic - a phenomenon that is also linked to the currently prevailing paradigms.



In reality, an evaluation-and-follow-up system will only be successful if the evaluations it produces are used,<sup>1</sup> though the information that the latter yield will not necessarily be utilized. As already mentioned, the fact that an evaluation is well designed and of a high quality does not guarantee that it will be used as a basis for decision-making, and hence many evaluations and studies end up sitting on the shelf.

Another difficulty is the lack of dissemination strategies. Some evaluations are very technical and not easily understood by those who might use them to make decisions, while others reach conclusions -and make recommendations- that are unfeasible or minimally relevant, given which we argue that the former must satisfy certain minimal requirements, such as the need to provide timely, relevant, high-quality information, if they are to be used, added to which various key role players -who “own” them, agree that they are useful and relevant, and promote their utilization- must take part in their design, implementation and dissemination.

### Education evaluation and the raising of quality

Given the above, it is clear that there is no direct relationship between educational evaluation and decision-making, and, hence, perhaps an even less direct relationship between evaluation and the raising of quality.

Evaluation may be formative, aiming to achieve improvement along the way based on its discoveries of what needs to be changed, or summative, setting out to make judgments about what has not worked and thus fostering accountability, but regardless of whether it is formative or summative, evaluation should not be seen as a panacea for all the evils and shortfalls of the education system, since, even though it can help to inform decision-making, suggest changes, raise educational quality and foster accountability, it does not suffice to bring about transformation.

### The use of educational evaluation in Mexico

In our country, the use of educational evaluation as part of endeavors to improve teaching-learning practices and formulate educational policy would appear to have been limited thus far. In this regard, the information yielded by the study entitled *El uso de los resultados de las evaluaciones del*

*INEE por las autoridades educativas* [The use of INEE evaluation results by the education authorities] (Silva, 2012), is very revealing, since the people involved in education who were interviewed for the said study expressed very divergent opinions about the three topics covered – namely, familiarity with the information produced by the INEE, the importance assigned to the said information, and the latter’s use in, and impact on, decision-making. Thus, for example, while most of the education officials interviewed at the state level knew that the INEE existed, only one fifth of them reported that they were familiar with the information that it produces, in marked contrast to those interviewed at the federal level, most of whom said that they were familiar with, and had access to, the said information. With regard to the use of, and value assigned to, the aforementioned information, the study reveals that the information produced by the INEE is predominantly used for discussion or consultation – what Silva (2012:101) refers to as “using the data to build a discourse”: Indeed, the said study (p. 102) concludes that the “INEE’s influence on the political process is limited” and also depends on how much people know about the evaluations, how well they understand them, and how they interpret them.

This limited use of the results of the different types of evaluation -not just of those carried out by the INEE, but also those administered by other authorities such as the Ministry of Public Education (Spanish acronym: SEP) and the state-level education authorities- is also referred to in one of the findings of the Critical Inventory Studies (Spanish: *Estudios de inventarios críticos*) of institutions that carry out educational evaluation at both the elementary and the lower-secondary levels, which were commissioned by the INEE in 2013. Indeed, in a series of interviews carried out as part of the aforesaid study of evaluation inventories in elementary education, various interviewees asserted that both federal- and state-level evaluations had little or no impact on the institutional planning and/or redesigning of policies and programs, and it was also found that, at both the elementary- and the lower-secondary levels, most local-evaluation results were only disseminated in the schools themselves, while the local authorities had no system pertaining to evaluation results.

In order to better inform decision-mak-

ing, evaluation must also be comprehensive enough to enable people to see the wood for the trees. However, this is no easy task, since educational evaluation in Mexico has not progressed at the same pace vis-à-vis the different components, processes and results of the National Education System (Spanish acronym: SEN), or in terms of the different evaluation types and methods that are used and their aims, being unable to see the wood because it has focused more on student and teacher components than on educational policies and programs, syllabi, methods, materials and schools.

Now that it is autonomous, the INEE is addressing the abovementioned concern by developing new evaluations, notwithstanding which the paving stones of the road leading to a comprehensive evaluation that take stock of the SEN’s different components are still being laid.

### The guidelines: a new tool

As already remarked, the INEE guidelines, which are recommendations aimed at informing decision-making about educational policy, establish specific goals and refer to key aspects of educational improvement based on the identification of the main problems, weaknesses or risks that must be addressed to fill a specific gap in the education system or remedy a particular shortfall in it.

Their validity is underpinned by evaluation findings, available tried and tested knowledge about the topic in question, past decisions that have proven to be relevant and effective, and designs legitimized via consultations with various key players to whom their contents have been divulged and their scope and viability confirmed.

Hence, two key aspects of the guideline-creation process that seek to build bridges between evaluation, the use of the latter, and improvement, are: (i) evaluation of the successes and shortfalls of government actions, and (ii) participatory creation.

### a) Evaluation of the successes and shortfalls of government actions

To create guidelines, it is very important to evaluate government actions – i.e. the policies and programs that the Mexican government has implemented in order address the problems in question. While the guidelines are policy recommendations that seek to influence decision-making, the scope of the

improvement is defined based on the analysis or weighing of the government actions taken, their shortfalls, and the challenges faced by them.

In this regard, the INEE has also endeavored to define a path or model for the evaluation of educational policies and programs that places more stress on the evaluation of policies than on that of isolated programs, given that tackling the various problems facing education usually involves more than a single program. For example, evaluating the success of government actions aimed at reducing dropout rates in lower-secondary schools means analyzing the design, implementation, results and impact of scholarship programs, early warning systems for detecting dropout risks, or of actions aimed at increasing parent and community participation and collaboration.

Hence, the first policy evaluation carried out by the INEE for the purpose of issuing guidelines set out to evaluate actions and programs for the education of children and youths from families of migrant day laborers, reviewing the successes and shortfalls of the educational-policy actions taken by the National Council for the Promotion of Education (Spanish acronym: Conafe), and the Program for Educational Inclusion and Equality (Spanish acronym: PIEE) of the Department of Indigenous Education (Spanish acronym: DGEI) and the National Institute for Adult Education (Spanish acronym: INEA), as well as studying the implementation of the aforesaid actions at the local level by means of selected case studies.

### b) Participatory creation

The starting point for the creation of the guidelines was also an awareness that even the best evidence available does not suffice to inform policy. As asserted by Weiss and other authors, in order to ensure that research findings form part of the government agenda and play a part in decision-making, we need to build the requisite consensuses. In other words, the guidelines should not only be based on high-quality, methodologically sound evaluations and research, but also have the biggest possible consensus and be adequately disseminated, which is why it is necessary to engage in dialogue with the different key protagonists.

In order for the guidelines that it issues to be legitimate and viable, and have a better chance of success, the INEE needs to

build consensuses, and thus one of the main prerequisites for the production of the said tools is that they be created in a participatory manner, which is why the model or path for the issuing of guidelines includes ongoing consultation with different members of society and educators as a crucial aspect of the said guidelines' creation. In view of this, both when drawing up its proposal for the creation of the guidelines aimed at improving the initial training of elementary-level teachers that were issued in September of 2015, and also in its proposal for the creation of the guidelines for improving policy relating to the education of children and youths from migrant day-laborers' families (which is still in the draft stage), among other people, the INEE consulted members of organized civil society, academics and government officials, who not only helped to define the educational problem that the guidelines were intended to address, but also collaborated in the search for alternative solutions to the said problem.

Via consultation, an effort has been made to anticipate the different legal, budgetary, technical, political and administrative limitations that could prevent the said guidelines from being accepted and hamper their implementation.

### Final comments

Given how difficult it is to turn evaluation into government action, the INEE has traced out a path for creating and issuing guidelines, in accordance with which, as well as gathering the relevant information yielded by the evaluations and studies that are produced by the itself and other entities -which help to diagnose the problem and propose possible solutions- the INEE carries out an evaluation of the public policies implemented, designed, precisely, to ascertain the successes and shortfalls of the said policies. It is at this stage that it is possible to encourage policy reassessment and the carrying out of other actions that had not previously been envisaged, and the INEE also brings together different protagonists to jointly analyze both the public problems for which a solution is being sought and also the key improvements that need to be made in order to solve the said problems.

Since the guidelines are a new tool, and an enormous opportunity, for linking evaluation to change, it is important to continuously reflect about how they are created

and, in due time, about possible changes to, or extensions of, the aforesaid tool. Another, equally big challenge has to do with the follow-up and updating that the INEE, jointly with NGO's, represented by the Social Advisory Council for Educational Evaluation and other participants who wish to join in, decides to carry out, heavily based on the actions and commitments of the education authorities at which the guidelines are aimed. Furthermore, the said path should be plotted out with the aim of supporting the INEE's main strategy of linking evaluation to use and educational improvement. €

- 1 MacKay. (2007). asserts that other important prerequisites in order for an evaluation system to be deemed successful are that there should be incentives for those who use the findings and sanctions for those who do not, and that users have the tools, training and funding that they need in order to make the suggested changes.

### References

- Mackay, K. (2007). "How to Build M&E Systems to Support Better Government". World Bank Independent Evaluation Group. Consulted on the 26th of January, 2016 at: <http://goo.gl/WoTird>.
- Miranda López, F., Martínez Bordón, A., Mendieta Melgar, G., (2013) "Inventario crítico de instituciones y organismos públicos y privados de evaluación educativa para Educación Básica y Media Superior". Study commissioned by the INEE.
- Moreles Vázquez, J. (2009). Uso de la investigación social y educativa: Recomendaciones para la agenda de investigación. Perfiles educativos, 31(124), 93-106. Consulted on the 26<sup>th</sup> of January, 2016, at: <http://goo.gl/DpL3dg>.
- Silva Laya, M. (2012). "Seguimiento de uso de resultados de evaluación del INEE por parte de autoridades educativas". Study commissioned by the INEE. From July, 2011 to July, 2012.
- Subirats H., J, Knoepfel, P., Larrue, C., Varone, F., (2008) Análisis y gestión de políticas públicas. Ariel, Ciencias Sociales. Spain.
- Weiss, C. H. (1998). Have We Learned Anything New About the Use of Evaluation? American Journal of Evaluation, N° 1, 21-33.

## 2016 Guideline-issuance Program\*

In 2016, the National Institute for the Evaluation of Education (Spanish acronym: INEE) will work on five projects aimed at drawing up guidelines for improving compulsory education.

### What considerations guide the drafting of the guidelines?

The drafting of the guidelines should be based on the participative identification of the educational problems to be addressed and the collaborative proposal of solutions to the said problems, with the aim being to ensure not only that the guidelines are legitimate, but also that they are feasible and inform decision-making that will help to raise the quality of education and make it more equitable.

### The criteria for choosing the 2016 guidelines are that:

1. evidence is available that makes it possible to identify educational problems, weaknesses or risks.
2. the problems identified must be tackled in the context of current educational policy
3. there exist educational evaluations or research whereby it is possible to ascertain the extent of the backwardness and its impact on educational quality and equity.

### What are the topics for 2016?

#### Project 1. Guidelines for improving the education available to the sons and daughters of migrant agricultural day laborers.

##### WHY ARE THEY NECESSARY?

The sons and daughters of migrant agricultural day laborers find it very hard to gain access to education because: (a) their families are continuously moving from farm to farm; (b) there are no educational services adapted to their conditions; and (c) they work as child laborers.

According to the 2009 National Survey of Day Laborers (Spanish acronym: ENJO 2009), the sons and daughters of migrant agricultural day laborers spend an average of 4.5 years in school, compared with the national average of 8.1 years, in addition to which around 78% of the said population is living in a state of multidimensional poverty

Although the Mexican State provides education to the aforesaid population through the National Council for the Promotion of Education (Spanish acronym: Conafe), the Elementary-Education Program for the Sons and Daughters of Migrant Agricultural Day Laborers (Spanish acronym: Pronim),<sup>1</sup> and the programs of the National Adult Education Institute (Spanish acronym: INEA),<sup>2</sup> nevertheless the aforesaid group has faded from the educational agenda, since, when the Pronim became part of the Program for Educational Inclusion and Equity (Spanish acronym: PIEE) in 2014, its budget was reduced to a mere 44% of what it had been in 2013.

Moreover, the evaluations of Conafe and Pronim that were carried out reveal recurrent problems – i.e. (i) high teacher turnover and a shortage of teachers possessing the specific profile needed to work with the group in question; (ii) scant integration and inter-sectoral linkage among the different programs; and (iii) minimal information about the characteristics and educational paths of the population in question.

#### Project 2. Guidelines for improving the education available to children and youths who live in indigenous communities.

##### WHY ARE THEY NECESSARY?

The Mexican State is heavily indebted to this population when it comes to the provision of education. The “Preliminary, Free, Informed Consultation of the Indigenous Peoples” (*Consulta Libre e Informada a Pueblos y Comunidades Indígenas*: INEE, 2014) reveals that these communities neither identify with the education that has been provided to them nor find it relevant.

According to the “Educational Overview of Mexico’s Indigenous Population” (*Panorama Educativo de la Población Indígena de México*: INEE, 2015, preliminary version), there are big shortfalls in infrastructure, material and teachers possessing the specific profile needed to work with the group in question, and 74.7% of all children and youths who speak indigenous languages are taught by people who do not speak the said languages, while speakers of indigenous languages aged 15 or over only spend a total 6.7 years in school.

Though there is a plethora of government programs aimed at the aforesaid population,<sup>3</sup> a preliminary review of the said programs, and of such evaluations as are available, reveals that they do not focus on the indigenous population, do not afford preferential attention to the latter, do not adapt to its conditions or characteristics, and are not interlinked.

The preliminary results of the Evaluation of Basic Teaching and Learning Conditions (Spanish acronym: ECEA; INEE, 2015) at the primary level confirm that there are big gaps between the indigenous schools and other schools in terms of the quality of their furniture, toilets, text books and other teaching materials, and the number specialized teaching staff.

According to the National Plan for the Evaluation of Learning (Spanish acronym: Planea: INEE 2015), 28.4% less indigenous students in the sixth grade of primary school reach achievement-level 1 in Language and Communication than do children in general public schools.

### Project 3. Guidelines for reducing dropout rates at the lower-secondary level.

#### WHY ARE THEY NECESSARY?

Mexico has one of the lowest percentages of youths between the ages of 15 and 19 enrolled in upper-secondary education, added to its high temporary and permanent dropout rates levels, which exceed those recorded for all the other school levels. In the 2013-2014 school year, the dropout rates for primary, lower-secondary and upper-secondary school were 0.6%, 4.7% and 13.1% respectively (SEP, 2014).

Although policies and programs aimed at reducing dropout rates were put in place when the Department of Upper-secondary Education was set up in 2005,<sup>4</sup> the problem persists, and the only two such programs to have been evaluated are:

1. the Scholarship Program, the evaluations of which make it clear that scholarships are not the best way of reducing dropouts, since the problem is more complex and other factors, such as the relevance of the education offered and school organization, need to be taken into account.
2. the *Construye T* (i.e. "Build Yourself") program, the evaluations of which identify many reasons why young people drop out of school, including boredom and lack of motivation.



**Project 4. Guidelines for improving the Tutoring Program (TP) for teachers who have just joined the Professional Teaching Service (Spanish acronym: SPD).**

**WHY ARE THEY NECESSARY?**

In this program's first year, the provision of tutoring has proven to be a highly complex endeavor. Up to October, 2015, 9,795 elementary-school teachers had enrolled in this program (i.e. only 45% of all those who had taken the diagnostic test at the end of their first year of service), and one of the main reasons for this low coverage rate is the fact that very few tutors have signed up to take the pertinent entrance examination.

The PT for new members of the SPD, which started in the 2014-2015 school year and was to be offered for two years by the local education authorities, was the first obligatory SPD support program set up for the purpose of professionalizing teachers.

Since the SPD is a recently designed system that has not been operating for very long, the PT has not been running for very long either and there are no evaluations to ascertain of progress. However, since both federal and state-level education officials all over Mexico report big delays in implementing these tutoring services, it is essential that we identify the bottlenecks in order to suggest timely ways to improve both the processes and the results.

**Project 5. Guidelines for improving the design and implementation of the ongoing-training program in the context of the SPD.**

**WHY ARE THEY NECESSARY?**

In Mexico the ongoing-training policy has been marked by: (i) a marked disconnect among the educational services and institutions responsible for initial and ongoing training and teacher updating; (ii) heavily fragmented teacher training programs; and (iii) scant detection and satisfaction of regional and local training needs (National Council for the Evaluation of Social Development Policy, 2010 and 2012).

In 2015, the National Office for the Coordination of the Professional Teaching Service (Spanish acronym: CNSPD) issued the National Strategy for Ongoing Training and Professional Development in Elementary Education, which plots out lines of action for the implementation of the training programs "to be taken by the staff evaluated in accordance with the law". In January 2016, it was announced that the SEP's departments of Elementary Education and Upper Secondary Education would be responsible for ongoing training.

Just like that pertaining to tutoring, this is a new policy that is still in the design stage, and hence it is imperative that, starting at the said stage, the INEE provide consultancy and help in the formulation of the ongoing-training proposal to be made in the context of the SPD, since the achievement of the aims set forth in the said proposal will, in large measure, depend on the latter's success.

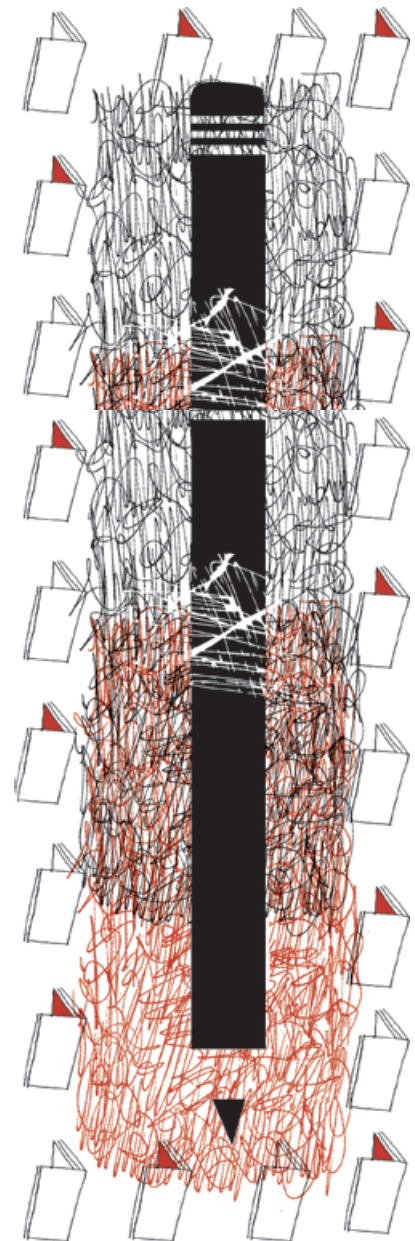
**What will the input be and what are the expectations for the 2016 guideline program?**

The input to the set of guidelines will be the results of the evaluations of the SPD in terms of learning outcomes, basic school conditions, policy analysis and assessments of government actions, the first two of which will make it possible to carry out a diagnosis of the situation to be addressed, while the last two will help to ascertain the successes and shortfalls of the actions taken by the government in order to solve the problem. €

\* Presented at the Ordinary Meeting of the INEE's Advisory Social Council on the Evaluation of Education (Spanish acronym: CONSCEE) on the 9<sup>th</sup> of February, 2016.

- 1 Currently a part of the Ministry of Public Education's Program for Educational Inclusion and Equity (Spanish acronym: PIEE).
- 2 Via the 10-14 Education-for-Life-and-Work Model (Spanish acronym: MEVyT 10-14).
- 3 Around 15 Federal Government programs have been set up to improve the education provided to the indigenous population.
- 4 The 2008 Comprehensive Reform of Upper-Secondary Education (Spanish acronym: RIEMS 2008) set out to make upper-secondary education relevant and facilitate the smooth passage of students between the different subsystems,

putting in place scholarship programs for impoverished students, risk-prevention strategies, tutoring and support systems, and remedial courses. In 2010, the *Síguelo. Caminemos juntos*, ("Keep Going; Let's Walk Together") program was established with the aim of linking previously unlinked efforts, followed, in 2013, by the *Yo no abandono* ("I Don't Drop Out") strategy, which included interventions at the academic, psycho-social and socioeconomic levels.



## ROADMAP

## Information systems on a par with the educational reform: innovation for strategic data management\*

Even though it is still in the development phase, one can already foresee the great strategic contribution that the Comprehensive Evaluation-results System (Spanish acronym: SIRE) -to be launched in the second half of 2016- will make to the working of the National Institute for the Evaluation of Education (Spanish acronym: INEE) and the National Educational-evaluation System (Spanish acronym: SNEE), and the impact that the said system will have on evidence-based decision-making and policy and syllabus design. The authors of the following article discuss the said sire and its probable effects.

**AGUSTÍN CASO RAPHAEL**

Head of the Unit for the Dissemination and Promotion of the Evaluation Culture of the INEE  
[acaso@inee.edu.mx](mailto:acaso@inee.edu.mx)

**ARTURO CERVANTES TREJO**

Director General of the Comprehensive Evaluation- Results System of the INEE  
[acervantes@inee.edu.mx](mailto:acervantes@inee.edu.mx)

**CECILIA GALAS TABOADA**

Deputy Director of the Comprehensive Evaluation-Results System of the INEE  
[cgalas@inee.edu.mx](mailto:cgalas@inee.edu.mx)

As a result of the Educational Reform and the setting up of the SNEE, we need to ensure that the different participants in the said system have access to reliable, relevant, high-quality information about the state of compulsory education in our country in

order to help evaluate the performance of the National Education System (Spanish acronym: SEN) and inform the taking of decisions aimed at improving it, and, in this regard, the plethora of data produced will be of great use in providing an overview of compulsory education in Mexico (Spanish acronym: SEP). Broadly conceived, as required by the Educational Reform, such information should facilitate decision-making, foster accountability, and promote transparency based on open-government practices.

An information system on a par with the Reform, and in line with the INEE's brief, should take stock of the needs of the strategic players in order to fulfill its purpose. Based on a user focus, as set forth in Figure 1 below, the said system should offer a selection of relevant, high-quality in a space that enables it to be nimbly administered and used while maximally exploiting the potential of the available data.

Developed by the Institute with the support of the Institute of Geography of the National Autonomous University of Mexico (Spanish acronym: UNAM), the SIRE brings together evaluation results and key information from the SNEE and the geographical, demographic and socioeconomic contexts, and, as a strategic project aimed at creating institutional value for the SNEE, provides an overview of education based on the ecological-systems model, bringing information from multiple sources together in a single platform for simultaneous use and, at all times, affording a geographical overview via geospatial analytical tools, as well allowing people to consult, analyze, mine and download data.

Given the SEN's large size and complex structure, the SIRE facilitates geographical analysis of the latter by indicating the location of schools, along with contextual information, and thus providing an overview that helps us to understand how the many factors that determine the real state of education in our country are interconnected. It depends, for its functioning, on a business-intelligence operating model in which the relevance of each piece of data is weighed in terms of the extent to which it enables institutions to achieve the aims of the Reform.



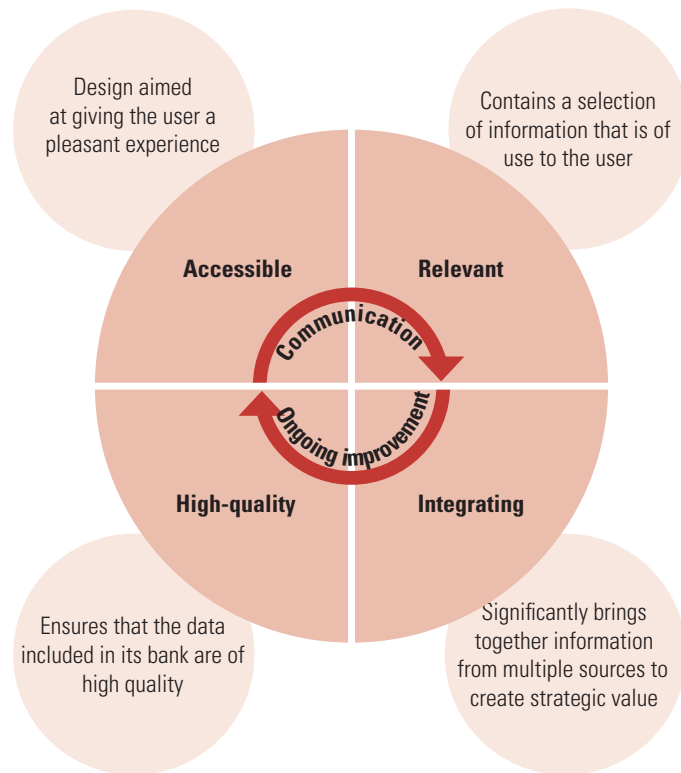
The said system, designed to satisfy the operational and data needs of the INEE and other Educational-evaluation System of the entities, takes stock of all the working and tactical information that stem from the Institute and the SNEE in order to choose, and bring together, all those strategic data that enable us to evaluate the Mexican education system's components and results.

Based on this focus on the value of information, and in line with the mission and aims of the inee and the National Educational-evaluation System, the sire compiles and manages a data bank that operates via the components shown in the following diagram.

- Data bank: data base for inputting the results of educational evaluations and strategically choosing key information about both the SEN and the geographical and socio-demographic context of education that are standardized to facilitate interlinked us.
- Geoportal: a technical tool for the carrying out of consultations and spatial analyses of a selection of evaluation-result data and data stored in other data bases pertaining to the system.
- Reports: documents statistics or systems available via Internet that compile key information, stemming from the data bank's contents, about topics having to do with the results of the evaluations of the SEN, and include access to contents, benchmarks, statistics and other educational-evaluation resources.
- Downloading of information sources: process whereby access is afforded to links, databases and files containing primary or processed information that can be downloaded from the system subject to legislation for the protection of personal data.
- Knowledge community revolving around the use of the system: forum for collaboration that enables progress and innovation in the education system to be divulged provides advice to users about the use of information and tools, and serves as a means of leveraging the SIRE's new capabilities.

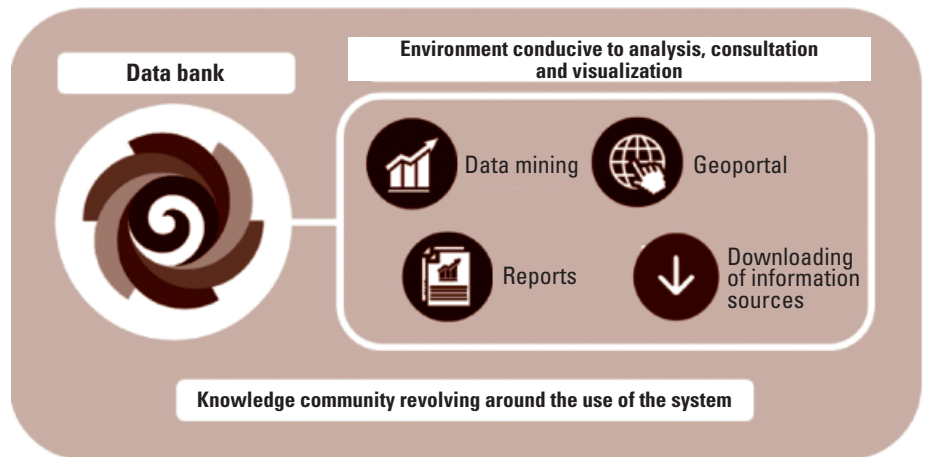
Among the SIRE's most outstanding features are its technological structure based on free-code software and its data model, which make it possible to operate jointly with other systems in order to exchange in-

Figure 1. Characteristics of strategic information systems



Source: Author-produced.

Figure 2. Components of the SIRE



Source: Author-produced based on the conceptual design of version 5.0 of the SIRE.



formation more safely and efficiently. The system's structure also makes it possible to access information from multiple sources and carry out consultation and analysis at different breakdown levels, whether geographical, or according to population features, type of educational service or user interests or needs.

Its versatility is a direct response to the need to make the information in question available to different audiences that play different roles in education and have different degrees of specialization. The applications and tools will be available -and useful- to high-level decision makers, state-level technical teams, researchers, members of NGO's, teachers, parents and students in accordance with the latter's characteristics and technical abilities.

Since the implementation of such an all-embracing system presupposes an institutional and cultural evolution towards knowledge-management models that support institutional strategies, and assumes the development of technical and human capacities for using highly advanced tools, the SIRE envisages the design of a change-

management strategy, and the creation of a knowledge community, in order to foster the adoption of a system that helps users to navigate the system in a simple, effective way, doing so at different levels according to their needs.

This SIRE was not only designed and developed based on the best international practices, but is also becoming a touchstone for the work being done in other countries and the top-level panels of entities such as the Organization for Economic Cooperation and Development (OECD) and UNESCO's State-of-the-World's-Education-Systems initiative.

At the national level, it is becoming a strategic tool enabling the INEE and the members of the National Educational-evaluation System to function, as well as a valuable strategic tool that enables the Institute and each of the SNEE's members to perform the tasks entrusted to them and supports decision-making, transparency and accountability in the area of educational evaluation. €

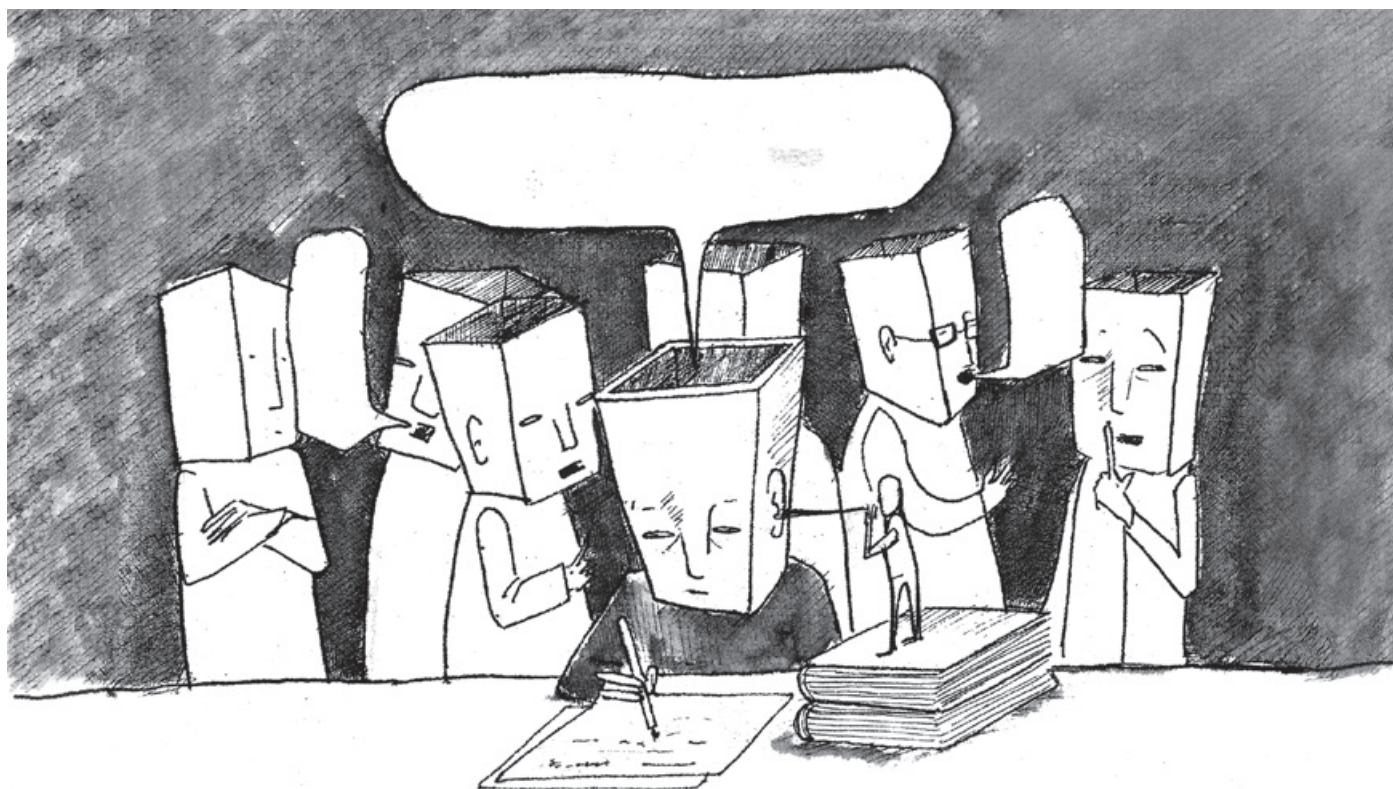
\* We thank Marisol Sánchez Barrera for her help in describing the sire's conceptual design.

## References

- Diario Oficial de la Federación. (11th of September, 2013). Ley General de Educación. Mexico.
- INEE. (2015). *Diseño conceptual del SIRE v5.0*. Dirección General del Sistema Integral de Resultados de las Evaluaciones. Mexico: INEE.
- INEE. (2013). *Estatuto Orgánico del Instituto Nacional para la Evaluación de la Educación*. Mexico. INEE.



Those wishing to find out more about the sire should visit the following link: <https://www.sire.igg.unam.mx/>



## IN THE CLASSROOM

## Can the project for certifying educational quality via ISO 9001-2000 improve education in Zacatecas?

Can high-quality education be achieved by means of inter-level linkage? Conceiving of supervision as a basis for educational decision-making, Adolfo Castruita Monreal analyzes educational policy in the state of Zacatecas, which is applying ISO 9001-2000 guidelines to elementary-education.

ADOLFO CASTRUITA MONREAL  
Primary-school-supervision unit number  
100 Ministry of Education of the State of  
Zacatecas  
[acastruitam@prodigy.net.mx](mailto:acastruitam@prodigy.net.mx)

### The ISO 9001-2000 standard in education

ISO standard IWA2-2003/NMX-CC-023-IMNC:2004, *quality-management systems: guidelines for the implementation of ISO 9001-2000 in education*, adopted as part of the International Workshop Agreements (IWA's) is deemed to be the general standard for the implementation of ISO certification in schools, since it enables us to analyze different aspects of the quality and management of state education.

Such certification sets out to promote ongoing improvement in the most common aspects of school operation and optimize learning outcomes by analyzing institutional needs, shortcomings and strengths. Oria Razo (2003) lists the guiding principles for implementing the ISO 9001-2000 standard in Mexican schools as:

1. A focus on getting to know the student and his/her educational needs, in order to provide fast, effective solutions that

exceed the latter's expectations, and even those of his/her parents.

2. Leadership on the part of the school principal, which implies wielding authority rationally and convincing people rather than giving orders.

The project aimed at certifying education in the state of Zacatecas using ISO international standards is a response to social and academic needs in a context where institutions are seeking to base their actions on parameters that enable them to measure the efficiency of public education services and detect problems in the different levels of educational administration.

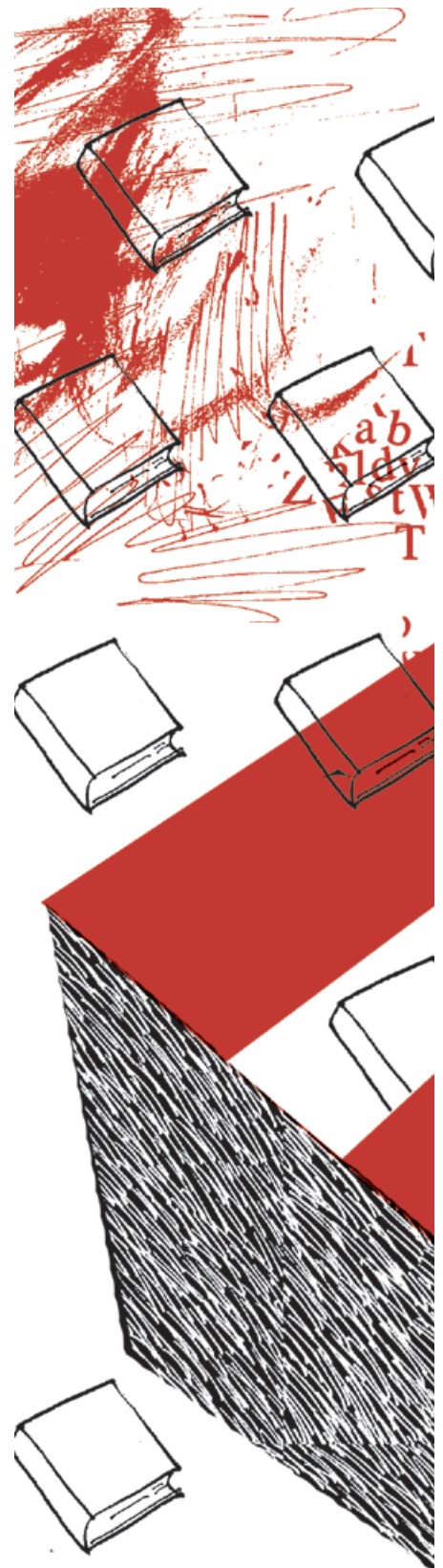
### Implementation of the ISO 9000 standard in Zacatecas

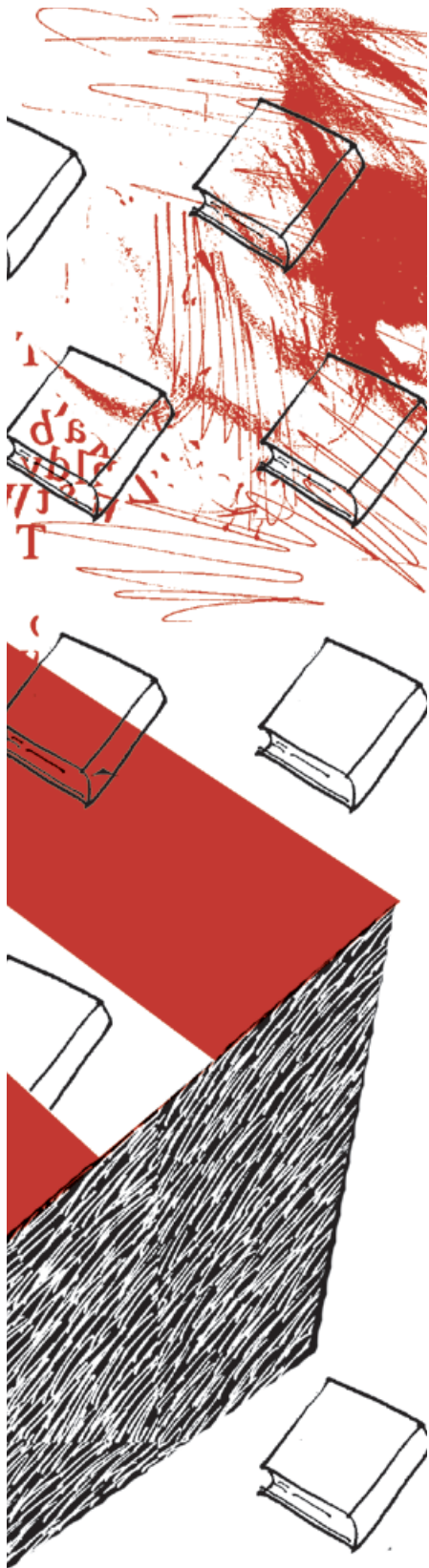
By reviewing and analyzing the implications of adopting ISO 9001-2000 in the state of Zacatecas for educational policy, it is hoped to significantly change the way schools are perceived by society and redefine what kind of schools the latter wants, since it may well be the case that changes in teaching methods, materials, classroom dynamics, as well as school equipment and maintenance, constitute part of the new educational model aimed at improving learning outcomes and thus enabling pupils to make sense of today's realities.

It is pertinent, for purposes of this project, that Zacatecas is implementing the High Quality Schools Program, with the State Council for Social Participation in Education cooperating in all the latter's stages, from training to the adjudication, selection and monitoring of each of the Strategic Plans pertaining to all the schools that have signed up for this general research project aimed at raising the quality of education in our state.

Given the above, the said Council decided that two elementary-level institutions participating in the High Quality Schools program -i.e. the Salvador Varela Reséndiz morning-shift primary school in the municipality of Zacatecas and the General Lázaro Cárdenas del Río general secondary school in the city of Fresnillo- should seek ISO 9001:2000 certification.

After just over one year and three months of joint efforts by teachers, manage-





ment and support staff and parents, in May, 2004, the aforesaid two schools obtained the said certification, which was endorsed by the International Certification Network (IQNet) company, a subsidiary of SAI Global Limited, so that the abovementioned Salvador Varela Reséndiz primary school, located in Laboristas street in the Zacatencas borough now has ISO 9001-2000 certification in management processes for high-quality education.

Following the guidelines set forth in the IWA2 governing document, the said school's pre-enrollment, enrollment, student-entry, teaching-learning, evaluation, certification and graduate-monitoring processes -all run in accordance with the standardized guidelines contained in the quality manual pertaining to the model proposed by the self-same school- were certified.

The aforesaid model seeks to reduce the number of shortfalls in the rendering of educational services, increase academic productivity, make optimal use of teaching time, increase commitment in pupils and their parents, and set up an ongoing-improvement process by making daily procedures more systematic and monitoring quality.

Despite the above efforts, the school in question is constantly criticized by -and often receives complaints from- some parents for various reasons such as rejecting requests for children to join certain groups, forming a first grade with so-called "recommended pupils", expelling children with low performance levels, having teachers who constantly make heavy demands on their pupils, and, sometimes, ill-humoredly scolding parents and pupils - so much so that, having reached the end of the program, it is hard put to gain recertification, and even runs the risk losing the certification it currently has.

The pre-enrollment, enrollment, student-entry, teaching-learning, evaluation, certification and graduate-monitoring processes of the General Lázaro Cárdenas del Río federal secondary school, located in Fresnillo, Zacatecas, are likewise certified, also being run in accordance with the standardized guidelines contained in the manual pertaining to the private-education quality model.

Via this certification, the said institution, which is one of the best schools in the municipality, taking in 400 students a year, and graduating a similar number who achieve very good results in subsequent educational levels, is endeavoring to raise the quality of the services it renders.

In order for a school to implement an ongoing-improvement program such as the one proposed here, it needs to make some changes to its internal organization in order to achieve more efficient teaching, and this implies the widespread use of different teaching methodologies, support materials and equipment in order to facilitate learning.

One of the priorities is that technological resources be used not only by pupils to increase their knowledge and acquire skills, but also by teachers and administrative staff, since the most important task is that of designing and producing teaching materials, especially when one considers that teachers are specialists in the subjects that they teach and should be capable of developing teaching methods that satisfy their pupils' needs.

As the research draws to an end, the General Lázaro Cárdenas del Río school is putting the finishing touches to its efforts to be recertified for a further three years, and it bears mentioning that the said institution's achievements are due to its principal's leadership and the clear commitment of its teachers.

### Conclusions

The lessons learned from this educational policy adopted in the state of Zacatecas in order to achieve significant changes in the current conception of what a school consists of and what type of school that society needs spring from the endeavor to ascertain the strengths and weaknesses of each school's teaching methods, practices and materials, equipment and upkeep from a private-sector viewpoint so as to formulate the pertinent ongoing-improvement plan.

Ideally, each elementary-level school should draw up its own proposal for improving learning outcomes, analyzing achievement during each school year, and proposing actions likely to result in ongoing improvement, since the aim is to raise pupils' achievement levels and equip them

with skills, as well as to augment their cultural capital, and endow them with the intellectual tools that they need in order to go on learning. €

### References

- Normas 9000. "Guía de inicio", consulted on November 30<sup>th</sup>, 2007. <http://www.normas9000.com/RegistrosRequeridosPorISO.htm>
- ISO 9000, ISO/IWA2/WD1N4 and Mexican Ministry of Education (Spanish acronym: SEP) standards (2002). *International Workshop Agreement 2. Quality Management Systems-Guidelines for the application of ISO 9001:2000 to education*.
- \_\_\_\_\_. The Ten Most Frequently asked questions about ISO, in *ISO 9001-2000, What is ISO and how does it work? Global Engineering documents*, consulted on November 15<sup>th</sup>, 2007 at <http://global.ihs.com/>
- ORIA Razo, Vicente. (2003). "La calidad de la educación", in *ISO 9001:2000 en la educación mexicana*. Ediciones virtuales, SEP, Mexico.
- \_\_\_\_\_. (2003). "Los ocho principios básicos para la aplicación de ISO/IWA2:2003, directrices para aplicar la norma ISO 9001-2000 en la educación", in *ISO 900-2000 en la educación mexicana*. Ediciones virtuales, SEP, Mexico.
- SCHMELKES, Sylvia. (1992). *Hacia una mejor calidad de nuestra escuela*. Publicaciones SEP.
- \_\_\_\_\_. (2000). "La Evaluación de los Aprendizajes en la Educación Básica", en *Educación 2001* No. 70 (March), Mexico.

## DOSSIER: USE OF INFORMATION, TWO PROPOSALS

### OTHER PERSPECTIVES

## Gazette No. 4. How to foster the use of evaluations?

**"The generally accepted tenet that evaluations do not lead to change due to the lack of interest of those in charge of programs or the scant evaluation culture does not suffice", say the authors of the following article, who give three reasons why evaluation results are not used to make decisions, four reasons why recommendations are not heeded, going on, finally, to describe two uses to which evaluations are put.**

GUILLERMO M. CEJUDO

Academic Secretary of the Centro de Investigación y Docencia Económicas (CIDE)  
[guillermo.cejudo@cide.edu](mailto:guillermo.cejudo@cide.edu)

FRANCISCO ABARCA GUZMÁN

Master's degree in Public Policy from the CIDE  
[fabaguz@gmail.com](mailto:fabaguz@gmail.com)

When a government official makes public-policy decisions, s/he has to take many factors into account, including budgetary constraints, political expediency, managerial ability, citizens' expectations and the available evidence. Such decisions are not usually taken by a single person, via an orderly, rational process where the aims are not in dispute, but, rather, there is seldom enough evidence available. What can we do, under such conditions, to ensure that the decisions taken result in improved public policies that are more likely to actually solve the problems that they address?

For several years now, one of the most frequent answers to the above question stresses the need to generate evaluations that provide information that that guides government decision-making, and hence institutions have been set up charged with evaluating policy,

and evaluations have been carried out -and, in the case of Mexico, specific mechanisms constructed- for the purpose of ascertaining the extent to which government heed evaluators' recommendations, notwithstanding which utilization remains a major challenge for this whole system of rules and processes that is founded on the premise that its evaluations will be used.

In this article, after giving a brief introduction to the problem, we present an analysis of how the federal government tends to use evaluations, and, based on the latter, make some proposals as to how their use can be fostered.

### The problem of use

With regard to social development, in our country an evaluation system that is compulsory, generalized, systematic and standardized has taken shape, and these features have become an international touchstone – above all for foreign governments that seek to ascertain how effective such interventions are, with a view to improving them. According to the National Council for the Evaluation of Social-development Policy (Spanish acronym: Coneval), based on the compulsory evaluations of federal budgeting programs, over 3,300 recommendations have been made to an average of 118 programs that were evaluated each year from 2008 to 2013. Although evaluation of our country's development policy has proved to be a very useful means of producing information for purposes of decision-making, whether about budgeting or regarding the redesign of programs, nevertheless one of the suppositions that is most prejudicial for the evaluation of government policy is the assumption that there is a direct link between the latter and the significant improvement of programs.

Such concern about using results exists in all evaluation systems and is explained by the difficulty of setting up mechanisms that coherently make explicit the cause-and-effect relationships among the information produced, the protagonists and processes involved, and the results expected of the evaluation. The commonly accepted assumption that evaluations do not lead to change due to lack of interest on the part of those responsible for the programs and en-

tities evaluated, or to the scantiness of the “evaluation culture”, does not suffice to explain their use at a broader level.

Regardless of the purpose for which an evaluation is carried out, its perceived success or failure tends to be based on the extent to which the information it yields is used. All evaluation endeavors should be conceived of in terms of how useful their findings are for improving the results of a given social intervention. The specialized literature on the subject has identified two main types of evaluation depending on the use that is to be made of the information produced. On the one hand, formative evaluation stresses feedback aimed at improving a policy’s end product (Weiss, 1998), while, on the other hand, summative evaluation enables us to determine the merit or value of a policy and make a “last judgment” about it (Russ-Eft and Preskill, 2001). Thus, the information produced by evaluations can help to redirect efforts and the commitment of resources, or to put an end to a set of actions that is not yielding results, and, in either case, use implies the taking of policy actions as result of a growing understanding of the causes of a public problem (Pawson, 2013; Cejudo and Maldonado, 2011).

There are various reasons why the results of an evaluation are not used to inform decision-making. The literature suggests reasons ranging from institutional and policy factors to the individual attributes of decision-makers (González, 2016). Generally speaking, the features and quality of the evaluation system, and the characteristics of the evaluators, are the main factors that can impede use.

1. *Features of the evaluation system.* The management, design and commissioning of the evaluation (done at the moment when the program is being designed) can lead to temporary flaws or flaws in the process (e.g. if a bad or ad hoc evaluator is hired).
2. *Evaluation quality.* If an evaluator carries out a shoddy analysis, his/her findings will probably be of little or no use, and if an evaluator’s conclusions are too general or abstract, useable recommendations will not be able to be made. If the said recommendations are not feasible

or there is nobody responsible for implementing them, they it will be unlikely to serve as input for decision-making.

3. *Features of the evaluatees.* Certain attributes of evaluations and the recipients of their results -i.e. the degree to which the evaluation is institutionalized and the coexistence of decision-making ability, managerial ability and spending power in the person/entity that receives the evaluation results- that can explain how and why the evaluations are used. It is matter of the ability of organizations to react to evaluations, including the ability

to involve the right agents to make real, and not merely operational, changes and the match between evaluation-system guidelines and the rest of the of the accountability systems. In short, it is a matter of whether the recipient of the evaluation results wants -and is able- to act on the evaluator’s recommendations.

#### How are evaluations used in Mexico?

Based on the results published by Coneval in its Follow-up Reports on Aspects of Federal Programs that are Subject to improvement, we can assert (after Cejudo and Abarca, 2016)





that those evaluation results that identify so called Aspects Open to Improvement (AOI's) are indeed used by the entities responsible for budgeting. Since 2008, most of the said AOI's are categorized by the aforesaid entities as "specific", meaning that, besides acknowledging the need to improve these aspects, the entities in question acknowledge that they are directly responsible for making the recommended improvements. Moreover, year after year, except for the period 2012-2013, those responsible for using the information have, in most cases, recognized the high priority of the best proposals stemming from evaluations. In this regard, there is evidence showing that the information produced by the evaluation system is well received by those in charge of programs, who see it as an important information source.

Nevertheless, the aforesaid use varies. For example, the programs with the biggest budgets make more use of information that enables them to improve aspects having to do with their running or implementation, while, on the contrary, the programs with the smallest budgets use the information stemming from the compulsory evaluations to improve their own design (Cejudo and Abarca, 2016). This is because the former programs have been able to carry out additional research projects and evaluations that provide them with more accurate diagnoses of the problems that they face, while, for the latter, the compulsory evaluations are the only source of information, and hence, for the said low-budget programs, the quality of the said information, and therefore the quality of the evaluations themselves, plays a basic role in improving their interventions in the public sphere (Abarca, 2014).

Besides understanding how evaluations are used, we need to ascertain why the recommendations stemming from them are passed over by those in charge of programs. It has been found, in the case of Mexico, that recommendations are not heeded mainly for the following four reasons (Cejudo and Abarca, 2016):

1. *Lack of synchronization.* The recommendations that are made had already been considered by those in charge of the program as a result of previous evaluations, meaning that there is a lag between the times when evaluations are carried out and their implementation.
2. *Limited empowerment.* The recommen-

dations cannot be heeded because they exceed the program's brief, which attests to ignorance on the part of the evaluator.

3. *Lack of clarity.* The recommendations are ambiguously expressed by the evaluators, or are vague to the evaluation format.
4. *Lack of resources.* The recommendations cannot be acted on due to lack of money, staff or time.

The evaluation system itself can take stock of all these aspects in order, in the first place, to improve the quality of the information offered to those in charge of the programs, and, in the second place, to increase the probability that the said information will be used. The following section of this article contains recommendations on how to achieve this.

### What can be done to increase evaluation use?

Perhaps the most important point is that it is hard to impose use – i.e. if a government official doesn't know what s/he can use evaluation results for, how they were produced, or what their relevance is, s/he is very unlikely to act on the recommendations in question. In this regard, deliberate interventions are needed in the following two areas:

#### a) Evaluation quality

- i) The processes for commissioning the evaluations need to be improved, keeping a balance between involving those responsible for making key decisions (so that they do not see the evaluation as something alien) and maintaining the independence of the evaluator (so that closeness to the decision makers does not jeopardize the impartiality of the findings).
- ii) The evaluation needs to be designed and developed in such a way that the information gathered, and the decisions documented, produce a cumulus of evidence.
- iii) It is necessary to promote the hiring of better evaluators, who have ample theoretical and working knowledge of the programs and combine methodological soundness with expertise about the workings of the public sector.
- iv) It is necessary to demand that the evaluators make precise, relevant, feasible, timely recommendations involving the

clear identification of those responsible for implementation, and this means that "quality control" is not limited to the due and timely delivery of information within the terms of reference, but implies interaction aimed at assuring that recommendations are heeded in accordance with the same terms.

#### b) Organizational ability to act on recommendations

- i) It is crucial to identify the specific people responsible for acting on the recommendations so as ensure that decision-making ability, managerial ability, and, ideally, the ability to make budget decisions, all coincide.
- ii) All those involved in decision-making, both inside and outside the organization, should be made aware of the recommendations and the implications of their participation.
- iii) Forums need to set up for interaction between the evaluator and the decision-makers, so that the government officials understand the reasons for the recommendations and link them to their basic activities. In other words, the officials should not see the heeding of the said recommendations as a mere bureaucratic formality, but rather as a decision, that helps them to get results.

#### Conclusions

To turn findings into recommendations, there must be a dialogue between the evaluator and the person in charge of the program, and this can only happen if the former has reliable information, has carried out a sound analysis, and makes timely, relevant recommendations on the one hand, and the latter, on the other hand, considers the evaluation to be necessary and useful, understands the recommendations and sees in them an opportunity to improve his/her performance and achieve the aims set for his/her organization. If the aforesaid conditions are not fulfilled, even if the program head chooses to comply with the recommendations as a matter of obligation, is forced to act on them, and produces follow-up reports, the evaluations will not be as useful as expected.

In short, rather than being a putative aim or an afterthought, the use of evaluations must be at the core of decision-making about policies and their assessment, so that,

the conditions needed for the said evaluations to engender decisions that improve policy results already exist when the recommendations are made. €

#### References

- Abarca, F. (2014). *Evaluar con múltiples propósitos: efectos del sistema de evaluación mexicano en los programas de desarrollo social*. Master's Degree Thesis. Mexico: CIDE.
- Cejudo, G. and Abarca, F. (2016). ¿Cómo se usan las evaluaciones? Un análisis de cómo los programas reaccionan a lo que recomiendan las evaluaciones. Mexico, unpublished.
- Cejudo, G. and Maldonado, C. (2011). *De las recomendaciones a las acciones*. Mexico: CLEAR-CIDE-SFP.
- González, D. (2016). *Usos de la información de Seguimiento y Evaluación (SYE) en las Políticas Públicas*. Mexico, doctoral thesis, CIDE.
- Consejo Nacional de Evaluación. (2009). *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de Programas Federales 2008*. Mexico.
- Consejo Nacional de Evaluación. (2010). *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de Programas Federales 2009*. Mexico.
- Consejo Nacional de Evaluación. (2011). *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de Programas Federales 2010*. Mexico.
- Consejo Nacional de Evaluación. (2012). *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de Programas Federales 2011*. Mexico.
- Consejo Nacional de Evaluación. (2013). *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de Programas Federales 2012*. Mexico.
- Consejo Nacional de Evaluación. (2014). *Informe de Seguimiento a los Aspectos Susceptibles de Mejora de Programas Federales 2013*. Mexico.
- Pawson, R. (2013). *The Science of Evaluation. A Realist Manifesto*. London: Sage.
- Russ-Eft, A. and Preskill, H. (2001). *Evaluation in Organizations*. New York: Basic Books.
- Weiss, C. (1998). *Evaluation*. New York: Prentice Hall.

## The Use Radar

Though evaluation is certainly essential in the context of the National Educational-evaluation Policy, it is no less crucial that its results be used to drive improvement. How do we decide how to use evaluation evidence and recommendations? What are the basic dimensions of evaluation use? Dr. Pérez Yarahuán posits four components -acknowledgement, quality, credibility and participation- that form the tips of a radar's crosshairs.

GABRIELA PÉREZ YARAHUÁN

Research professor attached to the *Centro de Investigación y Docencia Económicas* (CIDE); Research coordinator, CLEAR Center for Latin America.  
gabriela.perez@cide.edu

The ways in which knowledge is acquired have changed radically over the last 20 years in most parts of the world. Today, science possesses tools for obtaining data and specific information about nearly all the aspects of human life -biological, economic, cognitive and social- at a speed that was heretofore undreamt of, added to which processing capacity has increased substantially due to the improvement and large-scale availability of technology over recent decades. Sophisticated techniques for the quantitative and qualitative analysis of information have proliferated as a result of the ability of multiple members of an increasingly interconnected social and academic community to collaborate, and nowadays the scientific community has the tools needed to generate more evidence about -and explain- the factors that affect so many aspects of our lives.

However, the fact that such evidence is available is no guarantee that it will serve to bring about desirable social changes, since the possession of robust evidence does not suffice to ensure that knowledge is converted into the decisions needed to improve the lives of millions of people who suffer from fundamental privations.

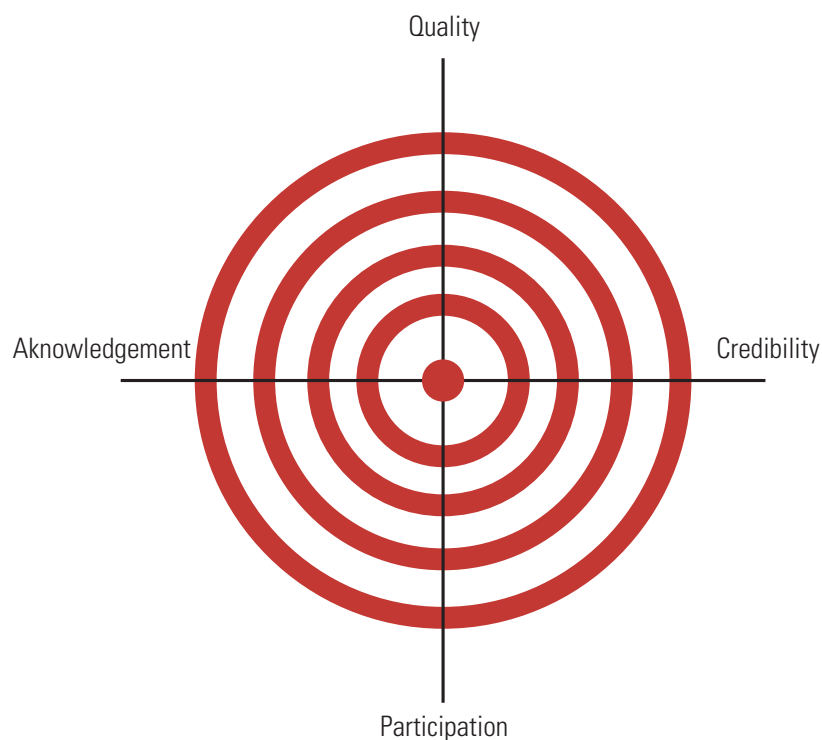
Without any doubt, in order to produce evidence that has an impact on the making of decisions about government policies or programs, we need rigorous scientific research that proves that given actions lead to the changes that are expected in the target population. However, it is not enough to determine, via research, whether a given circumstance is produced by a specific government action or program. Evidence is also a result of the analysis of processes of cause and effect and the identification of those factors that play a part in the logic of change. Since the causative chain (program theory) that stems from a desirable change must be analyzed in terms of the transformations that occur in people's behavior in response to an external factor or stimulus, and also of the proper implementation of the said stimulus, intervention or program by those responsible for taking action, then evidence consists in proving that government actions are effective, and providing logical programmatic explanations for this effectiveness and for the processes used to achieve it (Rossi, Lipsey & Freeman, 2004). In other words, evidence is basically gathered via a comprehensive process of evaluation that implies scientific ways of acquiring knowledge about the different aspects of the logic of government action and policy implementation.

However, the said evaluation will not be turned into evidence that is used for purposes of improvement unless certain features able to convert knowledge into action are explicitly considered. In this regard, the present article posits at least four dimensions, none of which suffices, in itself, to ensure the instrumental use of evaluation, though, together, they should inform decision-making about evaluation policy. The said dimensions are *acknowledgement, quality, credibility and the participation in the evaluation process of those responsible*





Figura 1. The Use Radar



for taking action, and the use of evaluation is the result of correct decisions in all four of them.

To clarify this idea, I posit a metaphorical "Use Radar"<sup>1</sup> (See Fig. 1) that should guide decision-making about evaluation policy.

Below, I explain the aforesaid four dimensions, how they relate to each other, why they affect the achievement of use, giving some examples to illustrate them, notwithstanding which the following two brief notes are required to clarify the concepts of evaluation and use.

#### **Type of evaluation and type of use**

The evaluation that is referred to in this article mainly consists of research into specific programs or interventions for the purpose of ascertaining if the actions taken (new

study curricula, teacher training, scholarships, provision of materials, building of infrastructure, expansion of school hours, etc.) result in changes in one aspect or another of people's lives (e.g. terminal efficiency, educational inclusion or student pupil performance).

When applied to evaluation, the term, *use*, has different meanings (Johnson, *et al.*, 2009; Leviton & Hughes, 1981). The evidence acquired via evaluation can be used to legitimize an action that has been carried out or learn about it, but it can also be used to make specific changes –e.g. in the type of goods or services provided, the inclusion criteria for the population served- or even as a justification for cancelling the intervention. When we talk about the use of evaluation to make improvements, we have in mind this type of instrumental use, which differs from



other kinds of use because it is hoped that specific actions, aimed at improving key aspects of interventions or programs, will ensue from the evaluation process. Some other types of use are symbolic use (or use in order to exert influence) and conceptual use (Herbert, 2014), which may set out to legitimize decisions or trigger a learning process, but do not result in specific actions regarding -or changes to- government programs.

All evaluations do not necessarily lead to instrumental changes, since not all evaluations are intended to have this sort of direct impact. For example, in the academic world, the purpose of evaluation-based research is to gather scientific evidence that helps to bear out hypotheses about factors that can affect human behavior (Levitt & List, 2009), with the hope that such an accumulation of evidence in a given sector will influence political debate and eventually lead to the formulation of more effective policies and greater accountability (Chelimsky, 2006).

However, the management-by-results model, which has flourished in government organizations, places special stress on evidence-based decision-making – i.e. it emphasizes a commitment to make decisions regarding budgeting and set up or close down programs based on the results of studies and evaluations, having a very strong focus on instrumental use.

For evaluation to be used instrumentally, the protagonists who make decisions about the design and implementation of government programs and policies need to be convinced that it is “useable”, and this will only be possible based on a set of evaluation-process attributes that includes acknowledgment, quality, credibility and participation, which are *sine qua non* dimensions for the design of evaluation processes or systems that permit a high degree of use.

### The acknowledgment of evaluation

To form part of routine government administration and be used to improve policies and programs, evaluation needs to be acknowledged and accepted by a group involved protagonists. By “acknowledged”, we not only mean included in the formal rules (laws, decrees, regulations, etc.) that are in place, but also perceived as a responsibility both by those in government and by citizens. This is essential for the building of an “evaluation culture” whereby the evaluation of government policies becomes an

ongoing requisite for the establishment of patterns of action and behavior.

Fort the last 10 years, there has been wide recognition in Latin America of the need to evaluate government programs and policies. A recent study of evaluation systems in the said region remarks that most of its countries have passed fundamental laws governing government actions that include obligations to evaluate the latter (Pérez Yarahuán & Maldonado Trujillo, at press). Nevertheless, although evaluation is widely acknowledged, its low level of utilization is repeatedly stressed.

Although it is essential that the State’s duty to evaluate be acknowledged if evaluation results are to form part of the formulation and implementation of better, more effective government policy, as the aforementioned study shows, acknowledging evaluation is not tantamount to using it.

### The importance of quality in evaluation

For its results to be used for purposes of improvement, evaluation must be based on reliable data, appropriate methodology, explicit hypotheses and solid theories. These attributes are indispensable requisites for endowing evaluation with quality, and when evaluation lacks the latter, it will be only sparsely used or run the serious risk of not improving action and hence not yielding good intervention results. Though high technical quality is a prerequisite, it does not guarantee that evaluation will be used to improve programs. One example of high-quality evaluation whose use for instrumental improvement is limited is the Education, Health and Nutrition Program (Spanish acronym: Progresá), which is currently called *Prospera*, having previously been named *Oportunidades*. Though there can be no doubt that the evaluation of this program, the findings of which started to circulate in 1999 (International Food Policy Research, 2001), has been very highly acknowledged for its high technical quality, it is equally true that the said evaluation, in its early stages, was used to legitimate and continue the said program, rather than to improve its design or implementation. While the intention of the evaluation was to enable *Progresá* to survive a change of government in a period of political uncertainty, it should not be denied that its use was circumscribed to achieving the pro-

gram’s continuation, rather than affecting its design and modifying its components in order to improve its results.

### Credibility as a prerequisite of use

Evaluation is more likely to be used for improvement if its findings are perceived by the relevant protagonists to stem from a totally unbiased process that includes adequate research techniques and practices and considers and synthesizes all the relevant information in the recommendations that arise from the data analyzed.

For evaluation to be highly credible, it must be perceived as part of a legitimate, transparent process based on adequate professional criteria, and it must also be clear, since its credibility depends on the transparency of its design, on the information that it uses, and on the independence of the evaluators.

The external evaluations of social programs that have been coordinated by the National Council for the Evaluation of Social-development Policy (Spanish acronym: Coneval) since 2007 are a very good example of credibility. Over the last decade, Coneval, which has played a key role in the creation of a legitimate, transparent evaluation system that uses high-quality information, has created or coordinated more than 1,000 evaluations, including, of course, evaluations of education programs and various programmatic processes thereof, such as design, impact and performance, and thanks to the said organization, we currently possess a significant body of information and data regarding key social-program benchmarks. However, to what extent has this evidence been used for improvement? The findings of a study of budgeting programs for elementary education from 2000 to 2014 indicate that there has generally been scant use of external evaluations of programs aimed at modifying the latter’s design via operating rules (Pérez Yarahuán, 2015). The factors most likely to promote changes in the said operating rules are evaluation design, the programs’ contextual variables, and the political environment.

So the credibility of evaluation is important, but its design plays a fundamental role in determining its use, due not only to the nature of what is being evaluated, but also because it is in the design stage that fundamental issues such as the role of the protagonists (evaluatees, evaluators and decision-makers) are determined.

### Participation

Though empirical studies of use have often found that both the quality of evaluation, and also its relevance, credibility and acknowledgement, affect its use (Johnson *et al.*, 2009), we nevertheless need to establish a more general framework for understanding change mechanisms. In this regard, it has been proposed that theories be formulated linking different types of use and different levels of analysis (individual, group and collective). In accordance with this viewpoint, the change brought about by an evaluation should be explained in terms of a mechanism that affects attitudes and actions and has the ultimate aim of creating social wellbeing (Mark & Henry, 2004).

The instrumental use of evaluation does not happen automatically after a recommendation has been made. Evaluation is done in specific contexts that affect the nature of its input, and, for it to have an impact on the design and implementation of programs, we need to put in place a mechanism for learning about, understanding and socializing its findings whose design includes the participation of those evaluated.

The four dimensions of our use radar -acknowledgement, quality, credibility and the participation- can provide orientation for an effective evaluation policy. In order for educational evaluation to be used to bring about improvement, we must recognize that high-quality evidence is essential...but then so are the evaluation's acknowledgement and credibility, and also the mechanisms of participation, appropriation and socialization. €

1 This concept is based on Diernier's "Trust Radar", 2011.

### References

- Chelimsky, E. (2006). The Purposes of Evaluation in a Democratic Society. In *The Sage handbook of Evaluation*. London: SAGE.
- Herbert, J. L. (2014). Researching Evaluation Influence: A Review of the Literature. *Evaluation Review*, 38(5), 388–419.
- International Food Policy Research. (2001). *Is PROGRESA Working? Summary of the Results of an Evaluation by IFPRI*. undated: IFPRI.
- Johnson, K., Lija O. Greenesid, Stacie A. Toal, Jean A. King, Frances Lawrenz & Boris Volkov. (2009). Research on Evaluation Use: A Review of the Empirical Literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410.
- Leviton, L. & Hughes, E. (1981). Research on the Utilization of Evaluation. *Evaluation Review*, 5(4), 525–548.
- Levitt, S. D. & List, J. A. (2009). Field Experiments in Economics: The Past, the Present, and the Future. *European Review*, 53(1), 1–18.
- Pérez Yarahuán, G. (2015). La influencia de la evaluación externa en los programas gubernamentales para la educación básica. *Revista Mexicana de Investigación Educativa*, XX (66), pp. 685–710.
- Pérez Yarahuán, G. & Maldonado Trujillo, C. (En prensa). *Panorama de los sistemas nacionales de monitoreo y evaluación en América Latina*. Mexico City: Centro de Investigación y Docencia Económicas/CLEAR.
- Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004). *Evaluation. A systematic Approach* (Seventh). Thousand Oaks: SAGE.



National Educational  
Evaluation  
Policy Gazette in  
Mexico

