Applying Dynamic Performance Management to detect behavioral distortions associated with the use of formal performance measurement systems in public schools: the case of Colombia

Carmine Bianchi (*), Robinson S. Salazar Rua

Department of Political Sciences and International Relations
University of Palermo
Via Ugo Antonio Amico 2-4, 90134, Palermo, Italy
Email: bianchi.carmine@gmail.com
Email: carmine.bianchi@unipa.it
Email: rssalazarr@gmail.com
*Corresponding author

Abstract

Test based accountability systems are widely used to evaluate the performance of public schools. These measurement systems are constantly exposed to behavioral distortions, which impact negatively the outcomes of education. The aim of this paper is to illustrate how “Dynamic Performance Management” (DPM) is able to support the design of consistent and comprehensive performance measurement systems that prevent behavioral distortions in the specific context of schools. The described approach allows one preventing, detecting and counteracting dysfunctional behaviors thanks to the systematic identification and understanding of how policy levers and performance drivers impact on the end-results (in terms of both output and outcomes), and how such results feed back on strategic resources. In order to frame the unintended effects generated by performance measurement systems introduced by the central government to evaluate public-funded organizations, we applied DPM to the case of Colombian district schools. The case illustrates how governmental measurement systems, focused on standardized tests, caused problems associated with narrowing curricula (i.e. reducing the scope of the content) to maximize test scores. An insight DPM simulation model describing such case-study is discussed in the last section of the paper.

Keywords

Dynamic performance management, Performance measurement, Human behavior, Behavioral distortions, Decision making, Case study, Public schools, Colombia.

Biographical notes

Carmine Bianchi is Professor of Business & Public Management at the University of Palermo (Italy), where is the scientific coordinator of CED4 System Dynamics Group. He is the director of the Doctoral level program in “Model Based Public Planning, Policy Design, and Management”, and of the Master level program in Public Management. Professor Bianchi is member of the Steering Committee of the “European Master in System Dynamics”, a joint degree among the Universities of Palermo, Nijmegen (the Netherlands), Bergen (Norway), and Lisbon (Portugal). This is an “Erasmus Mundus” funded program,
sponsored by the European Commission. He has published in numerous academic and professional journals. He also serves on the Scientific Committee of various academic publications. Professor Bianchi has an extensive international research and consulting experience with public and private sector organizations. Such activities consist in the design of policies and the outline of programs linking strategy and implementation. Consulting and education projects that Prof. Bianchi has conducted cover: strategic planning and control, performance management and reporting, as well as System Dynamics modeling for performance improvement and crisis prevention (Dynamic Performance Management).

Robinson S. Salazar Rua is a doctoral student in Model Based Public Planning, Policy Design, and Management at University of Palermo (Italy). He holds a Master degree in Business Administration at Radboud University (Netherlands), a Joint Master degree in System Dynamics at University of Bergen (Norway), New University of Lisbon (Portugal) and University of Palermo, and a Bachelor degree in Engineering at National University of Colombia. His current research interests are focused on Decision Making and Human Behavior.

1. INTRODUCTION

Public education is a subject of relevant interest for the main stakeholders in the society (i.e. government, schools, enterprises and families), due to its impact on cultural and socio-economic development. Public education costs are considerably high in governmental budgeting: for instance, New York State spent about $19,000 per student in 2011 (U.S. Census Bureau: Public Education Finances, 2011, Table 6). According to Boyne & Chen (2006), setting targets on performance indicators may lead to better outcomes.

This is a major reason for demanding more public education accountability. In order to foster such accountability, governments have adopted different methods and techniques, implying the use of performance measurement systems and assessment schemes based on standardized tests (e.g. SABER, NAEP, NAPLAN, and PISA). This approach enables governments to track a number of variables which are considered important to implement educational policies (e.g. graduation rate, dropout rate, test scores, school climate, socio-economic conditions, etc). These tools have increased the level of awareness about education in the entire social sphere in several countries of the world. They have also led to a growing concern on how to affect school performance in a sustainable way. In England, a set of policies has been developed and implemented in order to make public institutions more accountable. Such policies are based on the use of national standards and testing, school inspections and new measures. Huber et al (2007) state that these control mechanisms should take into account the importance of a multiple actor approach where diverse stakeholders (e.g. other schools, local college, and employer institution) interact to meet school outcomes. They also suggest the need to take into account
differences among schools, due to the social conditions of the community where they are located.

Most of these measurement systems are based on an output-oriented view, which implies a short-term focus. In many countries, as it is in the case of Colombia and even in the USA (Sexton et al., 2017), reward systems are aligned with the outputs of schools. Therefore, there are increasing expectations from different stakeholders towards the capability of academic institutions to meet the required standards.

The behavioral consequences of such methods on public schools at both an organizational and individual level have been illustrated by the literature (Hamilton et al., 2002; Pedulla et al., 2003; Koretz et al., 2001; Hout & Elliot, 2011; Bracey, 1989; Dorr-Bremme & Herman, 1986; Romberg et al., 1989; Stake, 1988). Such consequences affect school strategies, through decisions like narrowing the curricula, investing more time for exam preparation – rather than on learning oriented activities (e.g. student portfolios, multi-disciplinary projects, cultural activities, etc) – and focusing teaching efforts on only the best students.

These actions are due to the high pressure that governments have put on schools through test based accountability systems. According to Fullan (2006) and Hopkins (2007), these strategies have produced satisfactory results in the short run. However, in the long run they have not been able to sustain an improvement in high level competencies of students (OECD, 2001; Hargreaves, 2003; New Commission on the Skills of the American Workforce, 2007). In the USA, the act of Congress *No child left behind* institutionalized the score in standardized evaluations as the indicator to hold public schools accountable for their student achievements. At the beginning, there were positive expectations on the possibility to pursue the principles behind this act. However, it was later illustrated how the proposed approach based on standardized tests was not effectively contributing to improve the knowledge and skill level of students (Koretz & Barron, 1998; Neal & Schanzenbach, 2010; Cannell, 1987; Linn, Grave & Sanders, 1990; Shephard, 1990). It has been remarked that an increase in testing pressure has been associated with an increase in students’ dropout and failure rates, as well as with cheating scandals (Nichols, Glass & Berliner, 2005; Rich, 2012, 2013; Rich & Delaquérière, 2012; Winerip, Severson & Brown, 2013). Similar phenomena have been experienced in Colombia. In the last decades, the government has been interested in improving the educational quality of district schools, since their performance in national and international evaluations has recurrently been under the expected targets. Hence, the Colombian schools have been exposed to substantial changes by the execution of numerous reforms. That was one of the reasons why the Colombian government launched in 2015 a relatively new measurement system called “*Synthetic index of educational quality*”, to evaluate school performance through four dimensions (i.e. student progress, efficiency, improvement and school climate). Nevertheless, criticism has been raised with regard to the scope and limitation of such new metrics, which might lead to unintended
consequences on the educational quality and other outcomes of interest (Zambrano, 2015).

On the other hand, Finland, which is widely considered as the country with the best educational system in the world, does not use a system of standardized evaluations. Therefore, Finnish schools do not spend time and resources on test preparation (Sahlberg, 2006). Hence, a strategy based on learning, rather than measurement, is the main feature of Finnish education. In this context, teaching is seen as a commitment of multiple stakeholders, which prioritize lifelong learning, rather than training for increasing scores (Hargreaves et al., 2007).

The failure of these systems can be framed in the light of the literature on the unintended behavioral effects produced by formal performance measurement systems in organizations. Bianchi & Williams (2015) assert that performance measurement systems are likely to jeopardize service-delivery efficiency and effectiveness when programs focus on outputs and outcome targets without a proper understanding of the process and of the factors that determine them. Therefore, the design of performance measures that are based on cause and effect relationships has a significant potential to face an illusion of control, schizophrenic administrative behavior, performance paradoxes and bureaucratization.

The proposed approach in this paper implies an outcome and learning-oriented perspective that may prevent: (1) the use of only a long term view as a strategy to reduce uncertainty by predictability, (2) the adoption of only a short term view through an inversion of means with ends, and goal displacement (De Lancer Julnes, 2006), and (3) the implementation of dysfunctional strategies based on performance adaptation, which leads to a weak capability of the adopted indicators to distinguish between good and bad performers (Van Thiel & Leeuw, 2002; Leeuw, 2003; Meyer & O'Shaughnessy, 1993; Meyer & Gupta, 1994). This paper will outline a Dynamic Performance Management (DPM) framework, based on modeling and simulation, to support the design of consistent and comprehensive performance measurement systems in schools, through a feedback perspective.

An insight model has been developed by the authors of this paper to show how DPM can allow public school decision makers to frame how performance drivers impact on the end-results (in terms of both outputs and outcomes) and, in turn, how such results affect the accumulation and depletion processes of the strategic resources which primarily affect school performance.

A focus on performance outcomes and on their drivers, requires an understanding of the factors that influence the strategic resource endowment in such organizational systems, to pursue sustainable growth, organizational endurance and crisis resilience. This requires that a balanced set of strategic resources be built up and deployed. It also needs an inter-institutional view, implying that the organizations in a local area and the
strategic resources they share be identified. Therefore, to implement a DPM approach, an organizational perspective should be combined with an inter-institutional view. According to this view, school performance originates from the interplay of the policies carried out by several stakeholders in the educational system. This requires that school strategic plans strive towards the achievement of sustainable outcomes in the context where educational institutions belong.

Based on such perspective, in this paper we will question the capability of performance measures set by only external institutions (i.e. governments) with respect to schools, to support sustainable growth according to an outcome based view. In the particular context of Colombian district schools, we will discuss the unintended behavioral effects of such measures, and we will illustrate the risk of adopting a short-term view of performance. We will also propose an alternative approach that may combine the external parameters (i.e. benchmarks set by the government) with performance measures that reflect the challenges of a specific educational institution and its environment, to better pursue outcomes and sustainable growth.

2. PREVENTING BEHAVIORAL DISTORTIONS THROUGH OUTCOME ORIENTED PERFORMANCE MEASUREMENT SYSTEMS IN THE PUBLIC SECTOR

Performance measurement systems are never neutral with respect to the behavior of organizations and individuals, since they are designed to make people accountable on those same measures: they aim to affect actions towards a given direction. An inconsistent design of such systems – e.g. with respect to the task environment, to the system of rules which define decision makers’ roles in an institutional framework (Bianchi et al., 2010; Borgonovi, 1996), and to the outcomes of the affected actions – may generate unintended behavioral effects (Smith, 1993; Bianchi & Williams, 2015).

In particular, the specific dynamic complexity of today’s public sector may amplify the risk that the use of performance measures generates behavioral dysfunctions. The attempt to affect the behavior of organizations whose role is to deal with social ‘wicked’ problems (Bianchi, 2015; Head & Alford, 2013; Laegreid and Rykkja, 2014; Rittel & Webber, 1973, p. 160) – e.g. education, youth disadvantage, unemployment, social cohesion – should outline how a single agency is expected to contribute, through its actions, to the system outcomes. Research has emphasized how in today’s complex governance settings (Osborne, 2010; O’Toole & Meier, 2011) a single agency is able to manage only a subset of the strategic resources affecting the wider system’s outcomes. For instance, in order to improve the quality of life in an urban area, welfare organizations and schools may support families by acting on staff, skills, financial resources, networks, and administrative rules. The police may control crime by deploying staff, skill levels and equipment. Other agencies might provide better conditions to run business by acting on labor, urban infrastructure, parks, and public housing.
Performance measurement/management is a key to foster the implementation of a ‘whole of government’ approach (Boyle, 1999; Christensen & Lægreid, 2007, p. 33–35; Johnson, 2005; Pollitt, 2003), and to support an inter-institutional perspective (Bianchi, 2010, 2016) where policy coordination fosters better outcomes. In such perspective, setting the boundaries of the relevant system (beyond the short term and output oriented targets of each agency) may enable the process through which performance measures affect actions leading to outcomes.

The conceptual background of this analysis is provided by the literature on the dysfunctional behavioral effects produced by improper use of performance measures in the public sector. Regarding this, three major limitations of performance measurement have been identified: attribution, representation of quality, and goal displacement (De Lancer Julnes, 2006). Attribution implies a wrong, biased or partial causal connection between outputs and outcomes, and corresponding actions. Representation of quality is related to the use of a broad and detailed set of performance measures leading to unselective analysis and unclear interpretation of results (Bouckaert & Halligan, 2008, chap. 8; Van Dooren et al., 2010). Goal displacement is “a lack of goal congruence created by motivation to achieve some goals sought by the organization at the expense of other intended goals” (Flamholtz, 1996). An important cause of goal displacement is the inversion of means and ends. This occurs when performance measures and reward systems motivate people to achieve intermediate results rather than the overall organizational goals and outcomes (Smith, 1993, 1995). Another kind of important dysfunctional behavior is gaming, i.e. “a deliberate manipulation of behavior to secure strategic advantage” (Radnor, 2005; P. Smith, 1995, p. 298).

It is not the goal of this paper to conduct an extensive literature review on this topic. The focus of this article is, rather, to illustrate how – in the specific context of Colombian public schools – the use of only performance measures set by an external institution (such as the Ministry of Education) may generate unintended behavioral distortions.

In the Colombian case, illustrated in the second part of this paper, the previously mentioned behavioral distortions mostly occur when public schools try to reach target scores set by the educational authorities. This condition leads to unintended consequences inside schools, such as narrowing the teaching curricula and investing excessive training time to prepare the students to increase their scores in the standardized evaluations. This results into an inversion between means and ends, due to a misperception of outcomes (i.e. educational quality, attractiveness of the country, family efforts, holistic education, etc). Other dysfunctions, in this case, are associated with gaming the performance measurement system to receive socio-economic rewards.

3. DYNAMIC PERFORMANCE MANAGEMENT: A MODELING APPROACH TO DESIGN OUTCOME ORIENTED PERFORMANCE

---

1 For a more detailed literature review on the topic see: Bianchi & Williams, 2015, p. 403 - 407
MEASUREMENT SYSTEMS TO PREVENT BEHAVIORAL DISTORTIONS

Dynamic performance management (DPM) has been used in this paper as a modeling approach to frame weaknesses associated with inconsistent performance measurement systems (Bianchi, 2016; Bianchi & Williams, 2015) designed in Colombian educational institutions, and to support the design of a more robust set of performance measures.

By adopting an “instrumental” view of policy making, DPM challenges traditional, static and sectoral performance analysis by using a three layers framework (Figure 1). First, such systems should not focus only on the end-results, such as outcomes and outputs generated by implemented policies and actions in a given time period. They should also identify performance drivers i.e. the critical success factors for achieving these end-results. The dynamics of such measures should be continuously monitored and possibly improved, to influence the achievement of desired outputs and outcomes in the long run. To affect performance drivers, decision makers should identify the strategic resources to build up and deploy through their policies. Such resources are stocks of tangible or intangible assets available to pursue the performance targets. The endowment of such stocks is affected over time by the value of corresponding inflows and outflows, which can be either directly or indirectly influenced by decision makers. In the first case, explicit decisions (such as in the purchase of equipment, or staff hiring) will affect the change of strategic resources that can be obtained in the market. In the second case, those resources that cannot be procured from the market (such as in the case of population and many intangibles, like image) are affected through the outcomes generated by organizational performance. Second, DPM suggests that the boundaries of performance measurement systems should not be limited to a single agency, since an outcome based view of performance requires measures that can gauge how the organizations in a region will impact the final targets, possibly by working together as partners.
In the next section, a case study portraying risks of dysfunctional behaviors associated with the use of performance measures set by the ministry of education for Colombian public schools will be illustrated through the DPM approach.

4. USING DYNAMIC PERFORMANCE MANAGEMENT TO DETECT BEHAVIORAL DISTORTIONS IN COLOMBIAN PUBLIC SCHOOLS

4.1 CONTEXT

The Colombian Education Sector works in a decentralized way and its main objective is to insure the fundamental right of the education of the children and youth in the country. This system implies the existence of a sectoral organization where each governmental level has competences and complementary responsibilities to improve the delivery of the educational service. At the national level, the Colombian Ministry of Education develops policies and goals, establishes rules, regulates the educational service, and tracks evaluations. Furthermore, it provides School districts with technical and administrative assistance in order to strengthen their management capacity, and to allocate financial resources on education according to specific criteria. At the local level (i.e. Departments, Districts and Municipalities), the education secretaries manage the service by leading, planning, supervising and managing physical, human, and financial resources. They are also responsible for the results in terms of educational coverage and educational quality. Moreover, they provide technical assistance to the public institutions under their supervision, carry out teacher training, and apply rewards and sanctions to the district schools (Codesocial, 2009).
The Colombian Education System consists of four main stakeholders: families, schools, Government, and enterprises (Herrera Santana, 2007). The effective collaboration among such actors is meant to generate changes in the students, to face the new challenges in the modern society through education. Therefore, local agencies should perform their functions in a coordinated way to achieve important outcomes for the society. However, currently the educational actors are not aligned to achieve a shared vision towards a significant progress of Colombia. According to Llinas (1995), the education system is not consistent internally: its efforts are not being addressed towards the same path. First, the families are absent in the education process. Second, the link that should connect families, the society, government, and business in practice does not exist. Such deficiencies in the current system advocate the need for designing Colombian educational policies according to a causal, holistic and endogenous view, in order to grasp the inherent complexity of the system, instead of using a fragmented and additive approach.

In Colombia, two of the most important national evaluations are “SABER” and “ICFES”. The first one is designed to verify the progress in mathematics, language and sciences of the students who are in the primary and secondary education levels; the second one is designed to check the aptitudes of the students who are in the last grade of the Colombian Education System and who will possibly start their professional studies. Colombia has also performed international evaluations in public schools such as in the case of “PISA” and “TIMMS”. Such tests have been used to know the results of education from a socio-economic viewpoint, and they have tried to explain the factors that might have led to the obtained results (Cajiao, 2008). It must be noticed that the use of standardized evaluations is a relatively recent practice in Colombia. However, it has had a long history in countries such as the USA. Fernández Gómez (2005) asserts that this type of exams had a small impact on the decisions of the educators until recently. Nevertheless, in the United States, this condition changed dramatically in the year 2001 when the law No child left behind was passed. After this law was passed, the standardized evaluations had a significant relevance on the decisions of the government and schools. Currently, in Colombia a similar process is occurring. In fact, such evaluations are massively applied and their results are made public.

### 4.2. A FEEDBACK VIEW OF BEHAVIORAL DISTORTIONS ASSOCIATED WITH THE USE OF SCHOOL PERFORMANCE BENCHMARKING

Hamilton et al (2002) identified three important behavioral distortions that are caused by high pressure levels generated by governments on public schools to increase their standardized test scores: (1) narrowing the curricula to teach only contents that are evaluated in the exams, (2) increasing the training time to maximize test scores, and (3) increasing students’ abilities to pass tests by designing evaluations with a similar format to standardized tests. These dysfunctional strategies may lead to a score inflation effect, which implies an increase in students’ test scores without improving their aptitudes and skills in the areas where assessments are focused. The described phenomenon results into a
short-medium term increase in the output and into a long-term effect decrease in the outcome.

In the Colombian case, behavioral distortions are associated with the use of a reward system designed on a short term and output oriented pace. When public schools perform well in such tests, teachers receive bonuses and public recognition, and additional resources are granted to the school. From a preliminary field study (Salazar Rua, 2016) carried out, through semi-structured interviews, to understand the effects of standardized evaluations on the behavior of Colombian schools, it emerges that teachers perceive the evaluation system as a potential disadvantage for the schools located in poor areas. In fact, the evaluation system does not take into account how the socio-economic conditions (e.g. poverty, military conflicts, early pregnancies, etc) of the areas where schools operate affect school performance (Salazar Rua, 2016). In order to cope with this issue, schools have been inclined to increase their scores by allocating more teaching time to a narrow range of subjects, at the expense of others. As an unintended result of this unbalanced time allocation, the outcomes (i.e. educational quality, holistic education, family efforts, etc) have been dropping, in spite of slight increases in these scores. The risk of behavioral distortions, caused by such systems – as will be seen – requires the use of proper methods that may help government design “robust” school performance benchmarking frameworks.

Figure 2 shows, through a causal loop diagram, the short and long term effects associated with behavioral distortions in public schools due to perceived performance gaps when static and myopic performance measurement systems are used.

**Figure 2. Perverse effects of Governmental performance benchmarking systems on school policies.**

As it can be observed from figure 2, a perceived performance gap may change the logic for allocating teaching personnel in each school. “Performance gap” is the difference
between the desired and the current level of benchmarked performance. If such gap increases, pressure is put on school teachers to increase standardized test scores. In the short term, such efforts to quickly improve performance (e.g. by narrowing the curricula, focusing training on a narrow scope of subjects, etc) generate a score inflation effect (Hamilton et al, 2002), which leads to an increase in the scores obtained in evaluations without improving the outcomes of education (balancing loop “B1”). Such myopic strategy, originated by behavioral distortions due to perceived performance gaps, is antithetic to sustainable policies. Such policies imply an increase in the allocated teaching time to gain long-term outcomes in students’ skills. So, the higher such myopic efforts are, the lower the time allocated to pursue such outcomes will be (other conditions being equal). This would generate further performance reductions in the long run, leading to higher perceived performance gaps (reinforcing loop “R1”). The reinforcing loop “R2” illustrates how an increase in the performance gap generates a further reduction in the resources government allocates to such schools. This would generate lack of investments (e.g. improvement in classroom conditions, investments in educational materials, teachers’ training and hiring), that would further cause lower performance levels, and higher performance gaps in the long run.

4.3. USING DYNAMIC PERFORMANCE MANAGEMENT TO MODEL BEHAVIORAL DISTORTIONS ASSOCIATED WITH THE USE OF PERFORMANCE BENCHMARKING BY THE COLOMBIAN GOVERNMENT: PROBLEMS WITH NARROWING CURRICULA

A DPM approach has been used to model how governmental evaluation systems, focused on only standardized tests on a bounded range of subjects, led Colombian district schools to narrow the scope of teaching, which weakened students’ learning in the long run, as an outcome of a strategy aimed at maximizing test scores in the short run.

Figure 3 illustrates an instrumental view of DPM applied to the Colombian case. It allows one to identify relevant end-result measures to understand the impact of behavioral distortions on district schools in the short and long term. It also facilitates the identification of the performance drivers that influence the accumulation and depletion processes of strategic resources. By identifying such resources, it supports an assessment of the gaps between their current state and the external targets set by the government.
Figure 3. Short and long-term effects of myopic policies in allocating teaching time by Colombian public schools to increase test scores.

Figure 3 portrays how a change in test scores is affected over time by the dynamics of the following performance drivers: “% Time allocated to tested subjects” (i.e. Number of hours allocated to tested subjects over Total school hours) and “Balance and scope of learning” (i.e. Number of hours allocated in tested subjects over Time allocation benchmark). Such drivers describe a policy through which a reduction in the performance ratio (i.e. Number of tests above a target score divided by a benchmark target) leads to an increase in the time allocated to the tested subjects. When the performance ratio is lower than one, and the performance gap is significant, schools react by looking for an easy-to-reach and quick increase in test scores (e.g. by narrowing the curricula, or increasing time allocated to tested subjects). However, in the long term, the side effects of this policy negatively impact on the previous score gains. In fact, a higher fraction of time allocated to a bounded set of tested subjects (in proportion to the total teaching time) would worsen the balance and scope of learning. This would only generate a score inflation effect, rather than an improvement in students’ aptitudes and skills. In the long run, also for those students who have been initially successful in their test scores, such “superficial learning” would not be able to sustain further score gains; on the contrary, it would lead to a negative net change in the tests above the target score.

Figure 4 shows an ‘insight’ stock-and-flow simulation model, based on a DPM approach, which can be helpful to further detect the side effects of implementing school strategies focused on an output-oriented approach, without taking into consideration the long-term effects on the educational outcomes. The main purpose of this model is to
improve the understanding of complex dynamics, rather than mimicking the behavior of historic time series (Bianchi & Winch, 2006). Being an ‘insight’ model, the parameters it embodies and the results it generates through simulations should not be taken as accurate estimates. The model can be used, instead, to start developing an understanding by decision makers on how feedback system structure affects system behavior, i.e. performance. Therefore, it provides more details and corroborates the insights of the initial analysis based on the DPM chart in figure 3.

![Diagram](image)

Figure 4. An insight stock-and-flow model illustrating short and long-term effects of myopic policies in allocating teaching time by Colombian public schools to increase tests scores.

The stock-and-flow model illustrates how test based accountability systems might be a major cause of myopic policies, as it has been investigated in the case of schools (Hamilton et al, 2002) and, more specifically, in the Colombian context (Salazar Rua, 2016). The feedback structure in figure 4 emphasizes how the change in the teaching time allocated to tested subjects (narrowing of curricula) is a result of a two stages process. First, the perception of an unsatisfactory performance in the number of tests above target leads to an increase in the time allocated to tested subjects. This policy is adopted after a delay, which is due to a number of factors – such as the time to report and analyze information, and to make decisions. Second, the implementation of this policy, which is also affected by administrative delays, leads – in the short run – to an increase in the tests above target score. The model depicts the variable “Teaching time allocation to tested

---

2 Though the data included in this simplified model is deliberately disguised, the logic assumptions on which the model has been developed have been empirically tested (Salazar Rua, 2016).
subjects” as a stock, since it is a resource whose consistency affects service delivery. It also depicts the long-term effects of such resource, through a second-level performance driver (“Balance and scope of learning ratio”), in terms of a reduced capability of students to sustain the high scores gained in the past (score inflation effect).

Figure 5 portrays, through a DPM chart, the results from a simulation run of the insight stock-and-flow model based on the previously discussed myopic policies. The simulation has been run over a 10-years time horizon. This time extension has been set in order to capture the unintended side effects of such policy on the test scores and on students’ learning in the long run.

The model illustrates the case of a ‘generic’ school where the total number of tests (and students) is 500. Such school has been unsuccessful in the past in its students’ test scores: only 20% of them passed tests above the target score. In order to increase performance, and to pursue a 90% in the passed tests above the target score, the school starts to increase the time allocated to tested subjects.

As an effect of this policy, in the first three years the number of tests above target score per year increases, leading to a positive and rising net change in the affected stock of total tests (and students) above the target. It also decreases the level of low performing students. An improved school condition leads to an improvement in the performance ratio.

However, since the level of such ratio is still below the target (i.e. 1), the school continues to divert the available teaching time to the subjects on which test based evaluation is done. This gradually leads to a loss of holistic learning. This phenomenon is measured by the increase in the “Balance and scope of learning ratio”: the higher this ratio is, the higher the lack of holistic learning will be. As an outcome of this side effect, after the third year of simulation, the net change in the number of tests above target score per year progressively decreases, and becomes negative after the fifth year. One may not immediately detect such outcomes in the first three years from the graphs in figure 5. Such unintended outcomes are only latent until the fifth year. After that time, they become evident through a progressive decline in the stock of high performing students. The difficulty to perceive such phenomenon is associated with the feedback structure of the system, and namely to the delays affecting the two flows that lead to the net change in the tests above target score. It is also associated with the diminishing returns of additional teaching hours allocated to tested subjects. The understanding of such counterintuitive behavior portrayed by the dynamic complexity of the analyzed system is a benefit of using stock-and-flow simulation to enhance DPM. This benefit can be pursued if a learning facilitator is enabled to sketch and use the model through the participation of decision makers, and if the model is able to portray and to support the understanding of the dynamics of the end-results (outputs and outcomes), performance drivers, and strategic resources.
Figure 5. Results from a simulation run illustrating short and long-term effects of myopic policies in allocating school teaching time to increase tests scores.
5. CONCLUSIONS

As previously commented, *gaming the system* is a quite common distortion of both group and individual behavior in organizations. In order to face gaming behaviors in public organizations, and more specifically on public schools, the nature of the game must be changed (Lowe & Wilson, 2015). In this article we have shown how such phenomenon, leading to unintended outcomes, can be prevented – or at least counteracted – by the design and implementation of ‘robust’ and outcome-oriented performance management systems.

The model illustrated in the previous section of this article, shows how DPM can be used as an approach to challenge the quality and consistency of performance measurement/management systems, and possibly to suggest how to improve them, to deal with behavioral distortions caused by inaccurate design of performance measures in the specific context of test-based accountability systems in public schools.

The described approach can support governments and performance management system designers in public schools to adopt a behavioral and system dynamics perspective to assess the ‘robustness’ of the set of adopted measures. This implies an understanding of how policy levers and performance drivers impact on the end-results (in terms of both output and outcomes), and how such results feed back on strategic resources over time. Likewise, it is important to remark that the design and implementation of school performance management systems must take into account the contextual features for each public school, due to the fact that there are not two identical systems (Berman and McLaughlin, 1978; McLaughlin, 1990).

DPM has been used in this article to illustrate and to frame the unintended effects generated by test-based accountability systems adopted by the central government to evaluate public-funded organizations in Colombian district schools. The case illustrates how governmental measurement systems, focused on standardized tests, caused problems associated with narrowing curricula to maximize test scores. Since the main cause of the described myopic behavior can be related to the adoption of a tight set of performance measures (i.e. of indicators that are focused on a single subset of the problem), using the described model as a learning facilitation ‘vehicle’ might suggest a school to extend the boundaries (in both time and space) of the measures to adopt. For instance, this would mean to redesign performance management systems, by using also more long-term oriented measures (that may also try to gauge intangibles) than those that are used by the national government. Other outcomes from using this modeling approach could be related to the search of a constructive dialogue among schools, and between them and government, students, families, businesses, and other stakeholders in order to challenge – and possibly to improve – the quality and consistency of the adopted frameworks to assess educational performance.
Further research will be needed to investigate how DPM may support both national
governments and local school systems to design and use performance measures that prevent 
gaming and other behavioral distortions associated with policies through which institutions,
groups and individuals react to such measures to improve their own benefits. This is an
interdisciplinary field of studies in social sciences. As Campbell (1969) noted, “The more
any quantitative social indicator is used for social decision-making, the more subject it will
be to corruption pressures, and the more apt it will be to distort and corrupt the social
pressures it is intended to monitor”. Empirical evidence of this principle, known in the
literature as “the Campbell law” (Behn, 2011), has been found in different social fields, e.g.
ranging from policing to immigration, and climate change. Even education is a policy area
where social issues are deeply ingrained. Therefore, the attempt to apply any quantitative
indicator as an incentive to people for improving performance in such field runs the risk to
generate the unintended problems described by the Campbell law.

This risk provides a major reason for innovating the knowledge and practice of
performance measurement/management in such policy areas. DPM can provide a useful
method to advance in this endeavor.

The case and the insight model we have analyzed in the second section of this article can
be considered as an initial attempt to contribute in this direction. Further research, implying
the analysis of more cases and the development of more insight DPM models in different
geographic and social contexts, will be needed to this end.

REFERENCES

Available at
http://www.hks.harvard.edu/thebehnreport/All%20Issues/August2011.pdf

Berman, P. & M. McLaughlin (1978). Federal Programs Supporting Educational Change,
Vol. VIII, Implementing and Sustaining Innovations, R-1589/8-HEW, report
prepared for the U.S. Department of Health, Education, and Welfare, Rand
Corporation, Santa Monica, CA.

Sector through System Dynamics Modelling, From an ‘external’ to an ‘internal’
perspective, Special Issue of Systems Research and Behavioral Science, vol. 27.

Management through System Dynamics Modeling to Deal with Wicked Problems:
the Case of Societal Ageing. Systems Research & Behavioral Science, 32, (4) 502-505


Codesocial (2009). *Organización del sistema educativo: Conceptos generales de la educación preescolar, básica y media* [Organization of the Education System: General concepts of the pre-school, basic and vocational education]. Bogota, Colombia: Ministry of education


and Practice. Centre for Knowledge, Innovation, Technology & Enterprise (KITE), Newcastle University Business School, Newcastle upon Tyne, UK.


presented at the EGPA Conference, Bern.


Zambrano, M.F. (2015). *Alcances e inconsistencias del índice sintético de calidad educativa, diseñado y aplicado por el ministerio de educación de Colombia en el año 2015* [Scope and inconsistencies of the synthetic index of educational quality, designed and applied by the Ministry of Education of Colombia in 2015]. Bogota, Colombia: IQ Matrix foundation