




Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



Universitat Autònoma de Barcelona

PhD programme in Sociology

Department of Sociology

Faculty of Political Science and Sociology

The trajectories and enactment of accountability policies in education.

A focus on Chile

Doctoral Dissertation by

Lluís Parcerisa Marmi

Directed by

Dr. Antoni Verger Planells

Dra. Alejandra Falabella Ambrosio

Barcelona, June 2020

A la meva àvia Isabel.
Gràcies per la tendresa infinita i pel teu somriure incansable.

This research was supported by the Spanish Ministry of science, innovation and universities (former Ministry of Education, Culture and Sports) through a FPU (Formación de Profesorado Universitario) predoctoral scholarship.

Ref: FPU13/01090

ACKNOWLEDGEMENTS

I would like to show my gratitude to all the people that, directly or indirectly, have contributed and made possible this dissertation. Although the doctoral thesis is mainly written in English and Spanish, in the following lines, I am going to use Catalan, my mother tongue, as it is easier for me to express those words that come out directly from the heart.

En primer lloc, vull donar el meu sincer agraïment al meu director de tesi, l'Antoni Verger. Ha passat molt temps d'ençà d'aquella primera tutoria al teu despatx on vaig arribar empès per la Mar, amb l'objectiu de rebre el teu consell sobre el tema del TFG. També ha plogut molt des de l'inici d'aquelles primeres classes a l'assignatura de Política Educativa del Grau en Sociologia, i posteriorment al Màster en Política Social, Treball i Benestar. En aquest sentit, voldria agrair-te el suport intel·lectual i emocional que m'has brindat al llarg dels darrers anys, durant el desenvolupament del TFM i la tesi doctoral, així com també el teu compromís, humilitat i amabilitat. Al teu costat he après i espero seguir aprenent molt tant humanament com intel·lectualment. Gràcies pel teu mestratge, per la generositat, pels teus bons consells i per la honestedat i rigor intel·lectual. Gràcies també per transmetre'm la passió per la investigació i per aprendre. Per últim, et vull agrair la teva amistat sincera i que m'hagis recolzat i transmès calma, serenitat i optimisme, fins i tot en els moments personals més difícils.

Gràcies també a la meva codirectora, Alejandra Falabella, per les orientacions i pel suport durant les estades de recerca a Xile. Els teus consells m'han brindat una mirada local i crítica imprescindible per comprendre la realitat del país i varen ser crucials per a facilitar la meva immersió en el context xilè i per al desenvolupament del treball de camp.

Com a jove investigador, em considero molt afortunat per haver pogut desenvolupar la meva tesi doctoral en un espai tan estimulants intel·lectualment, i amb tanta calidesa humana, com el grup d'investigació en Globalització, Educació i Polítiques Socials (GEPS). Gràcies a totes i a tots els seus membres per ser una família, una escola de vida i un espai fratern, col·laboratiu i estimulants pel desenvolupament del pensament crític i per a la formació de joves investigadors i investigadores. Moltíssimes gràcies Pepe, Xavi B., Toni, Aina i Xavi R., per haver iniciat i posat els fonaments d'aquest projecte intel·lectual i col·lectiu per a l'anàlisi de les polítiques públiques i les desigualtats socials. Gràcies pel vostre exemple i per combinar magistralment el rigor científic amb un compromís inqüestionable amb la justícia social.

Mil gràcies també per la vostra amistat, pels bons moments compartits, pel suport mutu i per ensenyar-me tan: Adri, Alba, Alex, Andrea, Andreu, Anto, Camilla, Clara, Edgar, Gerard, Giulia, Judith, Laia, Laura, Marcel, Marjolein, Marta Comas, Marta Curran, Martí, Mauro, Natalie B., Natalie P., Òscar, Rosario, Sheila i Víctor. És un veritable privilegi participar en aquest espai amb totes i tots vosaltres. Malgrat no detallar cada petita cosa que us voldria agrair (i que sempre recordaré), us diré que inclús en aquells moments en que he estat més lluny (geogràficament), he pogut sentir el vostre escalf ben a prop.

També vull donar les gràcies a l'Aina Tarabini, per integrar-nos a mi i a l'Alex al GEPS i per convidar-nos a participar al projecte ABJOVES i l'avaluació del programa Magnet, i al Toni Verger per permetre'm formar part del projecte REFORMED (GA-680172), en el marc del qual he realitzat la meva tesi doctoral i on tan he après i gaudit treballant els darrers anys. La participació en els vostres projectes ha estat clau i ha contribuït a fer possible l'inici i la finalització d'aquesta etapa de formació predoctoral. Us estaré sempre agraït.

La major part d'aquest camí iniciàtic en el món de la sociologia l'he compartit amb l'Alex. Gràcies per la teva amistat i suport al llarg del Grau, el Màster i el Doctorat. Gràcies també per tots els moments viscuts a tots i totes les companyes del Grau en Sociologia i del Màster en Política Social de la UAB.

A nivell institucional, vull donar les gràcies al Ministeri de Ciència, Innovació i Universitats per finançar el meu projecte predoctoral mitjançant una beca FPU. D'altra banda, també voldria donar les gràcies a tot el professorat del Departament de Sociologia que ha contribuït a la meua formació acadèmica al llarg del Grau en Sociologia, del Màster en Política Social, Treball i Benestar, i del Doctorat en Sociologia i al personal d'administració i serveis de la casa, pel suport institucional en les múltiples gestions que he hagut de realitzar al llarg del doctorat.

La darrera estada de recerca a Xile es va poder dur a terme gràcies al suport econòmic del projecte H2020 INCASI (GA-691004), liderat pel Pedro López-Roldán i a qui agraeixo molt l'oportunitat de realitzar l'estada. També vull mostrar el meu sincer agraïment als grups de recerca, facultats i departaments (Graduate School of Education, Centro de Justicia Educacional, Centro de Investigación y Desarrollo de la Educación i Departamento de Trabajo Social de la UdeC), i especialment a les persones que m'han acollit en les diferents estades de recerca a la University of California Berkeley, Pontificia Universidad Católica de Chile, Universidad Alberto Hurtado i Universidad de Concepción. Gràcies al Rick Mintrop, Alejandro Carrasco, Alejandra Falabella, María Teresa Rojas, Gabriela Rubilar i a la Paulina Benítez.

El meu agraïment també a l'Adrián Zancajo (University of Glasgow) i en Cristóbal Villalobos (Pontificia Universidad Católica de Chile) per avalar aquesta tesi per a l'obtenció de la menció internacional.

La tesi doctoral no hagués estat possible sense la col·laboració voluntària i desinteressada de les persones participants en la recerca. A totes elles, moltes gràcies per regalar-me temps i coneixements imprescindibles pel desenvolupament de l'estudi.

Per últim, vull donar les gràcies a tots els companys i companyes amb qui he treballat, conviscut i/o compartit molts bons moments a Santiago de Xile, a San Francisco i a Berkeley. Ja per acabar, durant aquest trajecte evidentment ha estat vital el suport incondicional dels amics i amigues de Berga i Sabadell, de la meua família i molt especialment, de la meua companya de viatge, la Gisela. Has estat un pilar fonamental al llarg d'aquest camí. Gràcies per recolzar-me en tot moment, pels projectes compartits i especialment per la paciència amb les "anades" i "vingudes" constants de l'estranger.

El meu agraïment també per totes aquelles persones estimades que ens han deixat en els darrers anys: Vicente, Emilia, Filomena, Isa, Florin i a la meua àvia Isabel, que ens ha deixat en ple confinament per la pandèmia global. Gràcies pel vostre exemple. Us trobem a faltar.

Als meus nebots, Magí i Rita, gràcies per aportar-nos llum i alegria.

A la vida que vindrà.

A totes i tots vosaltres, mil gràcies sinceres per ser-hi sempre!

ABSTRACT

In recent decades, New Public Management (NPM) reforms have circulated globally in all kinds of policy sectors. The advocates of this reform model consider the public sector to be excessively rigid, bureaucratic and inefficient. NPM tries to modernize public sector's governance by incorporating a wide range of instruments, ideas, and organizational models from the private sector. In education, the three most disseminated components of this reform model include accountability instruments, learning standards, and decentralization. Accountability in education includes very variegated models and multiple design options. However, performance-based accountability (or PBA) is the accountability model that has spread further globally. PBA is inspired by principal-agent theory and is used to evaluating schools, school administrators, and teachers based on their students' performance on standardized tests. These school actors can face consequences (rewards, sanctions, external intervention, support, etc.) depending on their performance. Despite the wide dissemination of PBA, it is still unclear what are the reasons behind its adoption across countries with disparate governments, education systems, and levels of economic development. Likewise, the evidence on the effects of PBA is largely inconclusive or even mixed and, in fact, the literature on this topic points out the risk of accountability measures triggering non-desired effects such as cream-skimming, teaching for the test, or narrowing the curriculum.

This doctoral dissertation explores the trajectories and enactment of PBA reforms in different contexts with a particular focus on Chile. This country is considered a laboratory of neoliberal reforms and has a long history of experimentation with standardized tests and PBA systems. The objective of the investigation is two-fold. First, the thesis analyzes the rationales behind the adoption of PBA reforms and the factors that influence their recontextualization at the national level, with particular attention to the Chilean case. To this end, this study carries out a systematic literature review (n=158) on the trajectories of PBA policies at the international level, and a case study of PBA reforms in Chile (2006-2011) based on semi-structured interviews to stakeholders (n=27) and documentary analysis. Secondly, the thesis examines the implementation and effects of PBA reforms in different school contexts in Chile. The thesis combines a scoping review of the literature (n=150) on the implementation and effects of the PBA reforms, with the analysis of school responses based on a sequential mixed methods design approach, which includes a survey to the management team (n=200) and teachers (n=1130) in a representative sample of 79 schools, and 50 semi-structured interviews with school leaders and teachers. Finally, through a comparative case study that combines semi-structured interviews (n=7) with documentary analysis and surveys of managers and teachers (n=39), the thesis delves into the enactment of PBA policies and data-use in disadvantaged contexts.

The findings show that the trajectories of PBA reforms are not linear. Far from increasing convergence in the PBA models internationally adopted, the thesis highlights the central role of administrative regimes in explaining variation in the regulation and evolution of PBA in different national contexts. Second, the study shows that, in the case of Chile, the emergence of a double educational crisis caused by global and local drivers (e.g., learning crisis and students' protests) facilitated the opening of a window of political opportunity for the adoption of PBA policies. The political malleability of PBA - and the fact that it can be constructed as a solution to different diagnoses - is key to explain its selection as a policy solution to the educational crisis that Chile was going through. In terms of policy enactment, the thesis identifies five school responses to PBA regulations called induced alignment, de facto opting out, dilution, accommodation, and fabrication. The thesis shows that, in disadvantaged contexts, PBA policies can trigger short-term

improvement strategies that do not revert to a substantive pedagogical change. Furthermore, in those school contexts where negative perceptions about PBA predominate, these perceptions can inhibit the reactivity of objective performance pressures and data-use incentives.

In conclusion, the thesis highlights the relevance of meso-level variables to understand the adoption of PBA policies' trajectories. Finally, the thesis also shows how PBA policies can trigger different logics of action and the key role of meaning-making processes. In turn, the thesis challenges central premises of the PBA's theory of action and advocates to advance towards alternative models of accountability in education.

Table of contents	
ACKNOWLEDGEMENTS	4
ABSTRACT	6
INTRODUCTION.....	13
PBA as a core governance instrument in education.....	15
References.....	20
Appendix.....	26
Compendium of publications.....	27
CHAPTER 1.....	31
Reforming governance through policy instruments: how and to what extent standards, tests and accountability in education spread worldwide.....	31
Introduction.....	31
A ‘socio-historic’ approach to policy instruments	33
<i>Instrumentation: inquiring into the adoption of new policy instruments</i>	33
<i>On the evolutionary dimension of policy instruments</i>	35
Methods.....	36
Instruments’ choice: three rationales behind educational governance reforms.....	37
<i>NPM marketizers: adopting TBA to expand market competition and choice</i>	37
<i>Neo-Weberian states: governance instruments travel to continental Europe</i>	38
<i>Napoleonic states: the partial and uneven adoption of governance reforms</i>	40
New and changing uses: the evolution of NLSAs and TBA systems	41
<i>NLSAs: an ever-expanding instrument?</i>	41
<i>The ebb and flow of test-based accountability</i>	43
New constituencies: the emergence of economic and political subjects within educational governance reforms	46
<i>Economic interests in testing</i>	46
<i>Spreading like wildfire, but meeting firewalls</i>	47
Discussion and conclusions.....	50
<i>On the cumulative nature of educational governance reforms</i>	52
References.....	54
Appendix: List of terms included in the search protocol.....	63
CHAPTER 2.....	67
La Consolidación del Estado Evaluador a Través de Políticas de Rendición de Cuentas: Trayectoria, Producción y Tensiones en el Sistema Educativo Chileno.....	67
Introducción	67
La Rendición de Cuentas como Política Educativa Global: Aportes Teórico-Analíticos para Investigar su Adopción y Re-Contextualización	70
Estrategia Metodológica.....	73
La Producción de la Crisis y la Ventana de Oportunidad para la Reforma Educativa. 74	

El Estado Evaluador y la Rendición de Cuentas como Solución del Fallido Mercado Escolar Chileno.....	79
La Construcción del Sistema Nacional de Aseguramiento de la Calidad: Fase de Institucionalización y Disputas de la Política.....	84
Conclusiones	88
Referencias.....	91
CHAPTER 3.....	101
Rendición de cuentas y política educativa: una revisión de la evidencia internacional y futuros retos para la investigación.....	101
La emergencia de la rendición de cuentas basada en pruebas estandarizadas.....	104
Sistemas de rendición de cuentas basados en pruebas de aprendizaje: Múltiples opciones de diseño	107
Los múltiples efectos de la rendición de cuentas en educación	109
<i>El carácter mixto de los efectos de la rendición de cuentas</i>	109
<i>Efectos no deseados</i>	111
El proceso de adopción y de recontextualización de las políticas de rendición de cuentas	112
<i>El personal docente ante las políticas de rendición de cuentas</i>	113
El contexto social y económico de las reformas de rendición de cuentas.....	115
Discusión final: Cuatro premisas para futuras investigaciones.....	117
Apéndice	127
CHAPTER 4.....	130
In and out of the ‘pressure cooker’: Schools’ varying responses to accountability and datafication	130
Introduction.....	130
Context of the study: Governing schools through performativity and markets in Chile	132
Understanding the variation in school responses through enactment theory.....	133
<i>Interpretation as a key moment in policy enactment</i>	133
<i>The role of actors’ subjective perceptions of accountability pressures</i>	134
Methodology	136
<i>Data</i>	136
<i>Method</i>	137
<i>Quantitative research stage: Constructing school categories</i>	137
<i>Qualitative research stage: Characterizing school responses</i>	139
Findings: School responses to accountability regulations	141
<i>High pressure and pro-PBA: Induced alignment</i>	141
<i>Low pressure and pro-PBA: Easily accommodating accountability requests</i>	143
<i>Low pressure and con-PBA: Between dilution and opting out</i>	144

<i>High pressure and con-PBA: Fabricating better scores</i>	147
Conclusions	149
References.....	151
Appendix.....	154
CHAPTER 5.....	159
To align or not to align: The enactment of accountability and data-use in disadvantaged school contexts	159
Introduction.....	159
Theorizing schools' responses to high-stakes accountability: meaning-making and policy translations in context.....	162
The Chilean case: main features of the performative school market	163
Methods.....	165
<i>Sample, data and units of analysis</i>	165
<i>Data analysis</i>	168
Findings.....	168
School A: Low-reactivity to PBA regulations.....	168
<i>Principled resistance to PBA and standardized tests</i>	169
<i>Data use in context: coping with cultural and material contingencies</i>	171
<i>Test preparation as a symbolic practice</i>	172
School B: Aligning to PBA expectations with fidelity.....	174
<i>Trust in numbers: a game of love and hate towards PBA instruments</i>	174
<i>An expansive and intensive data use</i>	176
<i>Routinized practices of test preparation</i>	177
Conclusions	179
References.....	182
Appendix.....	190
SUMMARY OF MAIN RESEARCH FINDINGS: PLACING POLICY TRAJECTORIES AND ENACTMENT PROCESSES IN CONTEXT.....	198
Global policies, diverging trajectories: a political sociology of PBA reforms.....	198
Unpacking school responses to PBA policies in the Chilean education system: exploring the role of subjective variables in policy enactment processes	202
CONCLUSIONS.....	205
Methodological implications and limitations	207
Policy implications: the need for rethinking high-stakes accountability systems	208
Future lines of inquiry.....	210
References	211

INTRODUCTION

In recent decades, the global diffusion of New Public Management (NPM) and performance-based instruments (Gunter et al. 2016) has exacerbated the datafication of education and, at the same time, has transformed school governance, teaching practices and students' learning experiences (Jarke and Breiter 2020). The NPM reform package has been adopted in both developed and developing countries with the aim of increasing schools' effectiveness and education quality. NPM is linked to managerial ideas and policy principles (Ball, Junemann and Santori 2017) and has been advocated by both right-wing and left-wing governments and political parties, although for different purposes (Verger and Normand 2015).

NPM was adopted in the 1980s in countries such as the United States of America, Canada, and the United Kingdom, and is associated with the modernization and revolution of the public administration's governance (Hood 1995; Sehested 2002). During this period, neoliberal governments accused public services of inefficient (Barzelay 2001) and adopted NPM-informed reforms with the aim of 'debureaucratizing' public services (Verger 2007). According to Ball and Youdell (2008) and Ball (2010), NPM can be considered as a form of "covert or endogenous privatization". These forms of privatization are based on the premise that the public sector should emulate private corporations to become more efficient and improve its productivity. Nonetheless, Bellei (2015) questions the unequivocal identification of NPM with endogenous privatization, arguing that NPM instruments can be also adopted and enacted to benefit and strengthen public education. In fact, as will be seen below, NPM policies are polysemic in nature and can adopt various forms and instruments.

By implementing NPM, policymakers aim to introduce new governance mechanisms, instruments and techniques from the business world to improve the public sector's productivity and effectiveness (Common 1998; Vigoda 2003). The NPM reform package promotes outputs-based governance mechanisms, competition among educational services providers, decentralization and fragmentation of public services in smaller units, empowerment of clients and contractual relationships (Clarke, Gewirtz and McLaughlin 2000).

NPM often implies a changing role (not a retreat) of the state (Hudson 2007). NPM transforms the state's classical functions, strengthening its role as a "regulator, evaluator and distributor of incentives" among a dense and complex network of education services providers (Verger and Curran 2014, p. 256). In this sense, performance-based governance instruments are key components of NPM reforms and are crucial in understanding the rise of the so-called "Evaluative state" (Neave 1988; Maroy 2009). In the education sector, the NPM reform package is intrinsically linked to the Global Education Reform Movement (GERM), which has three main components or policy principles (see Table 1, in the appendix), namely standards, decentralization and performative accountability (Sahlberg 2016).

Firstly, the NPM agenda involves the definition of learning standards and a prescribed curriculum by national governments. In addition, there is a growing focus on core subjects (mathematics, science and reading), as well as a tendency toward narrowing the curriculum. This is partly due to the attention paid to international league tables on the part of *policymakers* (Breakspear 2012; Sellar, Thompson and Rutkowski 2017), and to the fact that the educational results in these subjects are perceived as a proxy to measure education quality (Sahlberg 2016).

The second component of this reform package refers to the decentralization of education and the importing of corporate management models into the public sector. The devolution of responsibilities to schools (e.g., school autonomy and school-based management models) and the emulation of managerial leadership models, aim to make schools "more business-like and more like businesses" since policymakers believe that private corporations are more efficient than the public sector (Ball and Youdell 2008; Ball et al. 2017, p. 3).

Third, "performative accountability" (Lingard, Sellar and Lewis 2017), also known as performance-based accountability (PBA), is considered a pivotal policy instrument of the global education reform movement (Ball et al. 2017; Sahlberg 2016) and it is one of the main components of the NPM reform agenda (Tolofari 2005). NPM-style reforms adopt PBA systems to govern schools, principals and teachers at a distance. PBA systems usually use data from national, census-based, large-scale assessments to classify schools and assign rewards and sanctions among educational providers and school actors. The assumption

behind PBA systems is that school actors need extrinsic incentives to increase their effectiveness. Thus, it is expected that by linking rewards and sanctions (for schools, principals, teachers and students) to learning outcomes, school performance and quality will improve (Mintrop and Ordenes 2017). To a great extent, the three policy principles of standards, decentralization and accountability are articulated through national, large-scale assessments, which allow governments to “steer” education providers “at a distance” (Osborne and Gaebler 1993).

PBA as a core governance instrument in education

Performative accountabilities (cf. Lingard, Sellar and Lewis 2017) and new “data infrastructures” (cf. Lingard, Sellar and Savage 2014; Sellar 2015) have been adopted and enacted in many countries around the world (Kamens and McNeely 2010; Lingard, Martino, Rezai-Rashti and Sellar 2015) with the aim of modernizing education governance and improving education quality and equity. International organizations (IOs) such as the World Bank, the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the Organization for Economic Cooperation and Development (OECD) have played a pivotal role in the dissemination of performance-based governance instruments, although they emphasize different configurations and policy models of educational accountability (Verger and Parcerisa 2018).

Generally speaking, accountability can be conceptualized as “a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences” (Bovens 2007, p. 450). In this sense, the literature on accountability has elaborated various well-established taxonomies to classify the different types and modalities of accountability (Anderson 2005, Darling-Hammond 2004; Leithwood and Earl 2000; Maroy and Voisin 2013; Ranson 2003; West et al. 2011; see Publication C in this dissertation). Nevertheless, despite the wide range of approaches to accountability (e.g., professional, market, social, managerial, etc.), administrative forms of accountability, linked to national large-scale assessments have had greater resonance among experts and decision-makers (Kamens and Benavot 2011).

In the educational field, performance-based governance instruments have been adopted in countries with educational systems as diverse as Norway (Camphuijsen, Møller, Skedsmo

2020), Canada (Jaafar and Earl 2008), the United States (Hursh 2005), Australia (Gable and Lingard 2016; Lingard and Sellar 2015), England (Ozga 2013), Germany (Thiel, Schweizer and Bellmann 2017), The Netherlands (Browes and Altinyelken 2020), Spain (Verger, Prieto, Pagès and Villamor 2020), Chile (Diaz Rios 2018; Falabella 2020), Brazil, Russia or China (Kauko, Takala and Rinne 2018), among others. Existing research points out that policymakers might engage with PBA instruments for political (Pollit and Bouckaert 2011), economic (Smith et al. 2004), ideational (Mehta 2013) and technical reasons (Lingard et al. 2016). Nonetheless, despite the broad international dissemination of PBA instruments, it is still unclear as to why policymakers with opposite political ideologies, and from countries with such variegated educational systems and levels of economic development, agree to adopt PBA instruments.

It is also still unclear how PBA instruments are enacted and evolve in different country-settings. Likewise, the regulation and implementation of PBA policies are highly complex (Honig 2006). As noted by McDermott (2007), there is often a gap between the intentions of policy design and its implementation. In the case of PBA policies, policy design is very relevant and according to Publication C, PBA can involve multiple design options and approaches that may affect how schools and individuals perceive and make sense of such data-intensive policy instruments. Beyond policy design, policy enactment processes are mediated by multiple organizational, social and economic contingencies (Braun, Ball and Maguire 2011; Spillane, Diamond, Burch, Hallet, Jjita and Zoltners 2002) such as, the characteristics of the local educational market and how each school is positioned in the hierarchy (Gewirtz 2003; Jabbar 2015; Lubienski 2006), socioeconomic composition (Lupton and Thrupp 2013), organizational culture (Gawlik 2015) or leadership style (Finnigan 2010), which may affect the way in which these policies are appropriated and translated by school actors.

From an academic point of view, despite the centrality of PBA instruments in the educational debate at national and international levels, we still have insufficient evidence in relation to how, why, for whom, and under what circumstances, accountability policies work. Thus, it is essential to generate new evidence to open the black box of PBA policies to better understand how material, organizational and ideational variables influence the recontextualization of PBA. In so doing, this dissertation aims to generate novel evidence

to gain a more nuanced understanding of the enactment of PBA policies and its effects in different school contexts.

Specifically, this thesis aims to delve into the trajectory(ies) and enactment of PBA reforms with a particular focus on the Chilean education system, which combines elements of administrative accountability with market accountability measures at its core (Falabella 2013). Chile has undergone important structural educational reforms since the 1980s, which have made it one of the most marketized education systems in the world (Bellei and Vanni, 2015; Cornejo et al. 2015; Falabella 2015; Sisto, Montecinos and Figueroa 2013). Chile stands out for having actively promoted competition between schools, through demand-based, competitive funding mechanisms (such as the well-known “voucher” scheme) and a highly liberalized school choice system (Verger, Fontdevila and Zancajo 2016). The main reason for choosing Chile as a case-study is that NPM instruments have been prominent in public administration reforms. Specifically, in this country, NPM has been a focal point of recent education reforms, carried out by both right-wing and left-wing government coalitions (Cornejo et al. 2015). Unlike other educational systems where PBA policies have been adopted to commodify the education system (as in the Spanish case), some investigations suggest that in the Chilean case, PBA instruments have primarily been adopted to regulate the school market and address some of the inequalities market dynamics generate (Falabella 2015; Verger, Bonal and Zancajo 2015).

In this sense, PBA instruments have been implemented with the aim of correcting structural “market failures” (Le Grand 1996; Verger and Bonal 2012), such as asymmetries of information for school choice among parents or non-desired practices, such as cream-skimming (Waslander, Pater and Van der Weide 2010), which have contributed to deepening social inequalities in the Chilean context, exacerbating the segmentation of school supply (Zancajo, Bonal and Verger 2014).

The present research seeks to examine the trajectories and the rationales behind the adoption of PBA in different administrative regimes, and also how these PBA policies are being adopted, regulated and enacted in the Chilean context in particular. In this sense, the doctoral thesis analyzes PBA reforms in education following a multi-scalar logic, and intends to explore two levels of analysis, namely policy trajectories and policy enactment,

with a specific focus on Chile. The main research questions of the dissertation are the following:

- RQ1: How and for what reason have PBA instruments been adopted in different administrative regimes? In which way have they been adopted in Chile?
- RQ2: Why have policymakers selected PBA as a suitable policy solution to address a range of educational problems in Chile? What role have drivers of an ideational and material nature, operating on different scales, played during the policy process?
- RQ3: What are the main effects of PBA policies in education? What are the key areas of agreement and disagreement in existing literature? What are the major mediating factors that influence the policy enactment of PBA?
- RQ4: How are PBA policies and related datafication practices enacted in different institutional and socio-economic school settings? How do subjective variables mediate policy enactment processes?
- RQ5: What role do subjective variables play during the policy enactment process in disadvantaged contexts? How are PBA policies interpreted, experienced and translated by school actors in disadvantaged contexts?

The dissertation is structured as follows: the first chapter (Publication A), on the basis of a systematic literature review (n=158), analyzes the policy trajectories of PBA on a global scale. Theoretically speaking, the chapter develops a socio-historic approach to policy instruments that combines insights from the political sociology of policy instruments and historical institutionalism. The investigation focuses on the instrumentation processes of PBA in a broad range of educational contexts and pays attention to administrative regimes as a key mediating variable to explain different policy trajectories and contextualizations of PBA. To explore the policy change and sedimentation of PBA instruments, the investigation analyzes layering and conversion processes through which policy instruments evolve and are institutionalized.

The second chapter (Publication B), based on documentary analysis and 27 semi-structured interviews with policymakers and stakeholders, explores the political economy of PBA reforms with a specific focus on Chile (2006-2011). Specifically, it draws on the “politics and semiotics of policy adoption” (Verger 2016) approach that combines insights from critical constructivism and cultural political economy (CPE). This approach emphasizes

the role of ideas during policy change processes but without omitting their interaction with political and economic interests, and institutional contingencies. In this regard, this research conceives the CPE evolutionary mechanisms, namely variation, selection, and retention (see Jessop 2010), as heuristic devices to explore the adoption of PBA reforms in Chile.

Subsequently, chapter 3 (which includes Publication C) presents a scoping review of the literature (n=150) relating to the enactment and effects of PBA in education. This publication conceptualizes the different models of accountability in education and maps the mediating factors in the enactment of PBA and its main effects in terms of educational outcomes and school actors' behavior.

Finally, chapters 4 and 5 (which correspond to Publications D and E respectively) provide detail with regard to the enactment of PBA instruments in Chilean schools through a mixed-methods research approach. Specifically, publication D relies on data from a large-scale survey to principals (n=200) and teachers (n=1130) of a representative sample of 79 urban schools from the metropolitan region of Santiago, Concepción and Valparaíso, which is complemented with 52 semi-structured interviews to school actors carried out in a sub-sample of 16 schools selected on the basis of a factor analysis. Theoretically speaking, this research draws on insights from policy enactment, sense-making, and decoupling theories, among others. Specifically, the study develops two indexes to explore empirically how school actors' subjectivity (namely perceived accountability pressure and alignment with the performative culture), in interaction with situated, material and external contexts, mediates school responses. Finally, publication E combines policy enactment and sense-making theories and examines meaning-making processes, and policy translations of PBA and performance data use. Based on the factor analysis mentioned above, Publication E carries out a more in-depth analysis of the enactment of accountability and data use in disadvantaged contexts. To this purpose, I selected two schools with a similar socio-economic composition but different subjective dispositions towards PBA as well as distinct levels of perceived pressure to carry out a comparative case study that combines semi-structured interviews with principals and teachers (n=7), surveys data from principals and teachers (n=39), and documentary analysis.

References

- Anderson, J. A. 2005. *Accountability in education*. Paris: UNESCO-International Institute for Educational Planning.
- Ball, S. J., Junemann, C., & Santori, D. (2017). *Edu. net: Globalisation and education policy mobility*. New York: Routledge.
- Ball, S. J., & Youdell, D. (2008). *Hidden privatisation in public education*. Brussels: Education International.
- Barzelay, M. (2001). *The new public management: Improving research and policy dialogue*. Berkeley: University of California Press.
- Bellei, C. (2015). *El gran experimento: Mercado y privatización de la educación chilena*. Santiago: LOM ediciones.
- Bellei, C. & Vanni, X. (2015). The Evolution of Educational Policy in Chile 1980-2014. In S. Schwartzman (Ed.), *Education in South America*. London: Bloomsbury.
- Bovens, M. (2007). Analysing and assessing accountability: a conceptual framework1. *European law journal*, 13(4), 447-468.
- Braun, A., Ball, S. J., Maguire, M., & Hoskins, K. (2011). Taking context seriously: Towards explaining policy enactments in the secondary school. *Discourse: Studies in the cultural politics of education*, 32(4), 585-596.
- Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance*. OECD Education Working Papers, 71. Paris: OECD Publishing.
- Browes, N., & K Altinyelken, H. (2019). The instrumentation of test-based accountability in the autonomous dutch system. *Journal of Education Policy*, 1-22.
- Camphuijsen, M. K., Møller, J., & Skedsmo, G. (2020). Test-based accountability in the Norwegian context: exploring drivers, expectations and strategies. *Journal of Education Policy*, 1-19.
- Clarke, J., Gewirtz, S., & McLaughlin, E. (Eds.). (2000). *New managerialism, new welfare?*. London: SAGE.

- Common, R. K. (1998). Convergence and transfer: a review of the globalisation of new public management. *International Journal of Public Sector Management*, 11(6), 440-450.
- Cornejo, R., Albornoz, N., Castañeda, L., Palacios, D., Etcheberrigaray, G., Fernández, R., Gómez, S., Hidalgo, F., & Lagos, J. I. (2015). Las prescripciones del trabajo docente en el nuevo marco regulatorio de políticas educativas en Chile. *Psicoperspectivas*, 14(2), 72-83.
- Darling-Hammond, L. 2004. Standards, accountability, and school reform. *The Teachers College Record*, 106(6), 1047-1085.
- Díaz Ríos, Claudia. 2018. "The role of national paradigms in the translation of test-based accountability". Paper presented at the 2018 Conference of the Comparative and International Education Society: Remapping Global Education. North-South Dialogue. Mexico, March 25-29.
- Falabella, A. (2013). *Accountability policy effects within school markets. A study in three Chilean municipalities* (Unpublished doctoral dissertation). London: Institute of Education, University of London.
- Falabella, A. (2015). El mercado escolar en Chile y el surgimiento de la Nueva Gestión Pública: El tejido de la política entre la dictadura neoliberal y los gobiernos de la centroizquierda (1979 a 2009). *Educação & Sociedade*, 36(132), 699-722.
- Falabella, A. (2020). The Seduction of Hyper-Surveillance: Standards, Testing, and Accountability. *Educational Administration Quarterly*, 1-30.
- Finnigan, K. S. (2010). Principal leadership and teacher motivation under high-stakes accountability policies. *Leadership and Policy in Schools*, 9(2), 161-189.
- Gable, A., & Lingard, B. (2016). NAPLAN data: A new policy assemblage and mode of governance in Australian schooling. *Policy studies*, 37(6), 568-582.
- Gawlik, M. A. (2015). Shared sense-making: How charter school leaders ascribe meaning to accountability. *Journal of Educational Administration*, 53(3), 393-415.
- Gewirtz, S. (2003). *The managerial school: Post-welfarism and social justice in education*. New York: Routledge.
- Honig, M. I. (Ed.). (2006). *New directions in education policy implementation: Confronting complexity*. Albany: State University of New York Press.

- Hood, C. (1995). The “new public management” in the 1980s: Variations on theme. *Accounting, organizations and society*, 20(2-3), 93-109.
- Hudson, C. (2007). Governing the governance of education: The state strikes back?. *European Educational Research Journal*, 6(3), 266-282.
- Hursh, D. (2005). The growth of high-stakes testing in the USA: accountability, markets and the decline in educational equality. *British Educational Research Journal*, 31(5), 605-622.
- Jabbar, H. (2015). How do school leaders respond to competition. *Education Research Alliance for New Orleans, Technical Report*.
- Jarke, J., & Breiter, A. (2020). *The Datafication of Education*. New York: Routledge.
- Kamens, D. H., & Benavot, A. (2011). National, regional and international learning assessments: Trends among developing countries, 1960–2009. *Globalisation, societies and education*, 9(2), 285-300.
- Kamens, D. H., & McNeely, C. L. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative education review*, 54(1), 5-25.
- Kauko, J., Takala, T., & Rinne, R. (2018). *Politics of quality in education: A comparative study of Brazil, China, and Russia* (p. 228). New York: Routledge.
- Le Grand, J. (2003). *Motivation, agency, and public policy: of knights and knaves, pawns and queens*. New York: Oxford University Press.
- Leithwood, K., & Earl, L. 2000. Educational accountability effects: An international perspective. *Peabody Journal of Education*, 75(4), 1-18.
- Lingard, B., Martino, W., Rezai-Rashti, G., & Sellar, S. (2015). *Globalizing educational accountabilities*. New York: Routledge.
- Lingard, B., & Sellar, S. (2013). ‘Catalyst data’: Perverse systemic effects of audit and accountability in Australian schooling. *Journal of Education Policy*, 28(5), 634-656.
- Lingard, B., Sellar, S., & Lewis, S. (2017). *Accountabilities in schools and school systems*. In G. Noblit (Ed.), *Oxford Research Encyclopaedia of Education* (pp. 1-28). New York: Oxford University Press.

- Lingard, B., Sellar, S., & Savage, G. C. (2014). Re-articulating social justice as equity in schooling policy: The effects of testing and data infrastructures. *British Journal of Sociology of Education*, 35(5), 710-730.
- Lubienski, C. (2006). Incentives for school diversification: Competition and promotional patterns in local education markets. *Journal of School Choice*, 1(2), 1-31.
- Lupton, R., & Thrupp, M. (2013). Headteachers' readings of and responses to disadvantaged contexts: evidence from English primary schools. *British Educational Research Journal*, 39(4), 769-788.
- Maroy, C. (2009). Convergences and hybridization of educational policies around 'postbureaucratic' models of regulation. *Compare*, 39(1), 71-84.
- Maroy, C.; Voisin, A. 2013. As transformações recentes das políticas de accountability na educação: desafios e incidências das ferramentas de ação pública. *Educação & Sociedade*, 34(124), 1-24.
- McDermott, K.A. (2007). "Expanding the Moral Community" or "Blaming the Victim"? The Politics of State Education Accountability Policy. *American Educational Research Journal*, 44(1), 77-111.
- Mintrop, R., & Ordenes, M. (2017). Teacher work motivation in the era of extrinsic incentives: Performance goals and pro-social commitments in the service of equity. *education policy analysis archives*, 25(44), 1-43.
- Osborne, D., & Gaebler, T. (1992) *Reinventing government*. Reading, MA: Addison-Wesley.
- Ozga, J. (2013). Accountability as a policy technology: accounting for education performance in Europe. *International Review of Administrative Sciences*, 79(2), 292-309.
- Pollitt, C., & Bouckaert, G. (2011). *Continuity and change in public policy and management*. Cheltenham: Edward Elgar Publishing.
- Ranson, S. 2003. Public accountability in the age of neo-liberal governance, *Journal of Education Policy*, 18(5), 459-480.
- Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In K. Mundy, A. Green, B. Lingard, & A. Verger (Eds.), *The handbook of global education policy* (pp. 128-144). West Sussex: Wiley-Blackwell.
- Sehested, K. (2002). How new public management reforms challenge the roles of professionals. *International Journal of Public Administration*, 25(12), 1513-1537.

- Sellar, S. (2015). Data infrastructure: A review of expanding accountability systems and large-scale assessments in education. *Discourse: Studies in the Cultural Politics of Education*, 36(5), 765-777.
- Sellar, S., Thompson, G., & Rutkowski, D. (2017). *The Global Education Race: Taking the Measure of PISA and International Testing*. Edmonton: Brush Education Inc.
- Sisto, V., Montecinos, C., & Figueroa, L. A. (2013). Disputas de significado e identidad: la construcción local del trabajo docente en el contexto de las políticas de evaluación e incentivo al desempeño en Chile. *Universitas Psychologica*, 12(1), 173-184.
- Spillane, J. P., Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy*, 16(5), 731-762.
- Thiel, C., Schweizer, S., & Bellmann, J. (2017). Rethinking Side Effects of Accountability in Education: Insights from a Multiple Methods Study in Four German School Systems. *education policy analysis archives*, 25(93).
- Tolofari, S. (2005). New public management and education. *Policy futures in education*, 3(1), 75-89.
- Verger, A. (2007). L'AGCS i la Política de l'Educació: Anàlisi dels Factors de Liberalització. Doctoral dissertation. Bellaterra: Universitat Autònoma de Barcelona. Available at: <http://ddd.uab.cat/pub/tesis/2007/tdx-0229108-170149/avp1de1.pdf>
- Verger, A. (2016). The global diffusion of education privatization: unpacking and theorizing policy adoption. In K. Mundy, A. Green, B. Lingard, & A. Verger, *Handbook of Global Education Policy* (pp. 64-80). John Wiley & Sons.
- Verger, A., & Bonal, X. (2012). La emergencia de las alianzas públicoprivado en la agenda educativa global: Nuevos retos para la investigación educativa. *Profesorado. Revista de Currículum y Formación de Profesorado*, 16(3), 11-29.
- Verger, A., Bonal, X., & Zancajo, A. (2016). What are the role and impact of public-private partnerships in education? A realist evaluation of the Chilean education quasi-market. *Comparative Education Review*, 60(2), 223-248.
- Verger, A., & Curran, M. (2014). New public management as a global education policy: its adoption and re-contextualization in a Southern European setting. *Critical studies in education*, 55(3), 253-271.

- Verger, A., Fontdevila, C., & Zancajo, A. (2016). *The privatization of education: A political economy of global education reform*. New York: Teachers College Press.
- Verger, A., Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: A political sociology of global education reforms. *Educational Review*, 71(1), 5-30.
- Verger, A., Prieto, M., Pagès, M., & Villamor, P. (2020). Common standards, different stakes: A comparative and multi-scalar analysis of accountability reforms in the Spanish education context. *European Educational Research Journal*, 19(2), 142-164.
- Verger, A., & Normand, R. (2015). Nueva gestión pública y educación: elementos teóricos y conceptuales para el estudio de un modelo de reforma educativa global. *Educação e Sociedade*, 36(132), 599-622.
- Verger, A., Prieto, M., Pagès, M., & Villamor, P. (2020). Common standards, different stakes: A comparative and multi-scalar analysis of accountability reforms in the Spanish education context. *European Educational Research Journal*, 19(2), 142-164.
- Verger, A., & Parcerisa, L. (2018). Test-based accountability and the rise of regulatory governance in education: A review of global drivers. In: A. Wilkins and A. Olmedo (eds.) *Education governance and social theory: Interdisciplinary approaches to research* (p. 139-158). London: Bloomsbury.
- Vigoda, E. (2003). New public management. *Encyclopedia of public administration and public policy*, 2, 812-816.
- Waslander, S., Pater, C., & Weide, M. (2010). *van der (2010). Markets in education. An analytical review of empirical research on market mechanisms in education* (Vol. 52). Paris: OECD Education Working Papers.
- West, A. P. Mattei & J. Roberts. 2011. Accountability and Sanctions in English Schools, *British Journal of Educational Studies*, 59(1), 41-62.
- Finnigan, K.S. 2010. Principal Leadership and Teacher Motivation under High-Stakes Accountability Policies, *Leadership and Policy in Schools*, 9(2), 161-189.
- Zancajo, A., Bonal, X., & Verger, A. (2014). Mercados educativos y segmentación de la oferta escolar: efectos sobre las desigualdades educativas en Chile. *Témpora: Revista de historia y sociología de la educación*, 17, 11-30.

Appendix

Table 1. The role of national assessments within the GERM

GERM principle	Definition and main policies	Role of national assessments	
Standards	Prescription of a national curriculum and establishment of quality standards	National assessments used to make sure schools meet and adhere to evaluable learning standards	
Decentralisation	Transfer of competences and authority from the central government to lower administrative levels	National assessments used to control state, regional, provincial and local authorities	
	Devolution of managerial and/or pedagogical responsibilities to principals and schools	National assessments used to govern at a distance a range of autonomous providers through the principles of outcomes-based management	
Accountability	Educational actors made responsible for their actions/results through some form of evaluation linked to consequences.	<i>Administrative accountability</i>	Test results attached to incentives or sanctions for schools, principals and teachers
		<i>Market accountability</i>	Test results used to inform school choice and promote school competition

Source: Verger, Parcerisa and Fontdevila (2019, p. 5).

Compendium of publications

Below, there are the articles and book chapters that are part of the compendium of publications. As can be seen, the thesis follows a multi-scale logic. While the first two publications focus on the policy trajectories and adoption of PBA reforms, the last three focus on policy enactment at the school level:

- **Publication A:** Verger, A., Fontdevila, C., & Parcerisa, L. (2019). Reforming governance through policy instruments: How and to what extent standards, tests and accountability in education spread worldwide. *Discourse: Studies in the Cultural Politics of Education*, 40(2), 248-270. <https://doi.org/10.1080/01596306.2019.1569882>
- **Publication B:** Parcerisa, L., & Falabella, A. (2017). La consolidación del Estado evaluador a través de políticas de rendición de cuentas: trayectoria, producción y tensiones en el sistema educativo chileno [*The consolidation of the evaluative state through accountability policies: Trajectory, enactment and tensions in the Chilean education system*]. *Education Policy Analysis Archives*, (25), 1-24. <http://dx.doi.org/10.14507/epaa.25.3177>
- **Publication C:** Parcerisa, L., & Verger, A. (2016). Rendición de cuentas y política educativa: Una revisión de la evidencia internacional y futuros retos para la investigación [*Accountability and education policy: A review of international evidence and future research challenges*]. *Profesorado. Revista de Currículum y Formación de Profesorado*, 20(3), 15-51. <https://recyt.fecyt.es/index.php/profesorado/article/view/54589/33220>
- **Publication D:** Verger, A., Ferrer-Esteban, G., & Parcerisa, L. (2020, forthcoming). In and out of the ‘pressure cooker’: Schools’ varying responses to accountability and datafication. In: S. Grek, C. Maroy & A. Verger (eds.) *World Yearbook in Education 2021: Accountability and datafication in education*. New York: Routledge.
- **Publication E:** Parcerisa, L. (2020, under review). To align or not to align: The enactment of accountability and data-use in disadvantaged school contexts. *Educational Assessment, Evaluation and Accountability*.

POLICY TRAJECTORIES AND ADOPTION

Publication A: Verger, A., Fontdevila, C., & Parcerisa, L. (2019). Reforming governance through policy instruments: How and to what extent standards, tests and accountability in education spread worldwide. *Discourse: Studies in the Cultural Politics of Education*, 40(2), 248-270. <https://doi.org/10.1080/01596306.2019.1569882>

CHAPTER 1

Reforming governance through policy instruments: how and to what extent standards, tests and accountability in education spread worldwide

Introduction

New forms of regulatory governance strongly rely on data-intensive policy instruments. These ‘new’ policy instruments are adopted on the top of more traditional fiscal and legislative instruments in an attempt of steering increasingly fragmented and multi-layered policy systems more effectively (Wilkins & Olmedo, 2019). Data-intensive policy instruments are technically complex and sophisticated in design. Their main functions are to collect new forms of information on public sector performance, and to manage public services’ conduct at a distance (Scott, 2000). In education, regulatory governance has meant the adoption of national large-scale assessments, test-based accountability, and explicit learning standards.

National large-scale assessments (NLSAs), which usually rely on the external evaluation of students’ learning through standardized tests, are the governance instrument that has spread fastest in education systems recently. In the last two decades, the number of NLSAs being enacted globally has expanded exponentially, with their presence being especially important in OECD and middle-income countries (Ramirez, Schoffer, & Meyer, 2018; Verger, Parcerisa, & Fontdevila, 2018). Furthermore, NLSAs, beyond a data collection device, have become an intrinsic component of test-based accountability systems (TBA). The assemblage of NLSAs, standards and accountability constitutes ‘a coherent and effective political dispositi^P’ (cf. Ball, Junemann, & Santori 2017, p. 4) that, according to international data sources, is being increasingly enacted to monitor teachers’ performance and promote competitive pressures among schools (Teltemann & Jude, 2018; Verger et al., 2018). In countries with TBA systems in place, school actors face consequences of a different nature (material, reputational, individual, collective, etc.) according to their levels of performance and adhesion to centrally-defined learning standards.

The origin of regulatory governance in education dates back to the 1980s, when mainly Anglo-Saxon countries introduced structural reforms in public administration following

the tenets of neoliberalism. These reforms attempted to introduce marketization and privatization in the delivery of public services and to this purpose governments, among other measures, started publishing school rankings on the basis of NLSAs results. Later on, with the intensification of economic globalization, other countries started adopting similar policies as a way to monitor and strengthen the competitiveness of their educational systems. New governance instruments have often travelled as part of broader education reform packages that also promote decentralization, school autonomy and the diversification of school provision. Concepts such as the Global Education Reform Movement (Sahlberg, 2016) or New Public Management reforms in education (Gunter, Grimaldi, Hall, & Serpieri, 2016) are often used in education literature to capture this international phenomenon.

In this paper, we argue that, despite their globalizing dimension and neoliberal origins, the reception and evolution of data-intensive governance instruments needs to be seen as context-sensitive, contingent and path-dependent (Kauko, Rinne, & Takala, 2018; Maroy, Pons, & Dupuy, 2017). Policy instruments such as NLSAs and TBA have been globally adopted, but the uses (and the intensity of the uses) given to these instruments are arguably contingent to the specificities of the political and institutional settings where they are embedded. Governments might adopt NLSAs and TBA for different reasons, and it cannot be taken for granted that these instruments are chosen to promote similar policy changes, or that end up deepening market dynamics in education wherever they are enacted.

To test these premises, we have systematically reviewed a corpus of 158 papers focusing on the political economy of educational governance reforms in different world locations. We analyze our data through the lenses of a political sociology approach to policy instruments, which we combine with analytical premises deriving from historical institutionalism. In this respect, we expect that the politico-administrative regimes to which countries adhere strategically mediate the variegated adoption and evolution of NLSAs and TBA in education. The paper is structured as follows. In the first two sections, we introduce our theoretical and methodological framework. In the sections that follow, we present our results according to the main analytical axes of our approach: instruments choice, the evolution of instruments' uses, and the new constituencies and subjectivities

generated by governance instruments. In the last section of the paper, we discuss our findings and conclude.

A ‘socio-historic’ approach to policy instruments

The emerging demand for global skills in increasingly inter-dependent economies, the challenges generated by technological innovation, and the comparisons of educational systems deriving from international large-scale assessments are contributing to the expansion of similar education reforms globally (Verger et al., 2018). To a great extent, standardized evaluations and TBA have become central instruments of an education reform approach that situates learning achievement as a key driver of national success in an increasingly competitive economic environment (Rizvi & Lingard, 2010).

In this paper we acknowledge the importance of international competition, among other global drivers, for the spread of educational governance instruments. Nonetheless, we adopt a political sociology approach to policy instruments, in combination with a historical institutionalism premise, to more explicitly capture how meso-level factors strategically combine with global drivers in the production of more complex and multi-scalar education policy landscapes. This approach is well-suited to observe and systematize the diverging policy trajectories that global reforms and global policy models follow. To this purpose, we structure our theoretical framework according to two critical moments of educational policy change, first, the adoption of new policy instruments and, second, the evolving uses assigned to these instruments once they are being enacted.

Instrumentation: inquiring into the adoption of new policy instruments

Policy instruments are central to both conceptualize and understand current public sector reforms and changing forms of governance (Le Galès, 2010). Although the adoption of new policy instruments has been conventionally conceptualized as a second order change (cf. Hall, 1993), or as a change of a mainly technical nature, policy instruments choice is a very political moment that does not always follow technocratic and pragmatic logics.

Instrumental choice is a moment with major potential implications for the future development of public systems. Many policy instruments create their own structures of opportunity in ways that were unforeseen when first adopted, and can generate broader political effects in governance structures and even in the main goals that policy systems are

expected to pursue (Bezes, 2007; Kassim & Le Galès, 2010). For these reasons, political sociologists such as Lascoumes and Le Galès (2007) invite us to problematize the logic of instruments choice and conduct further empirical research on what they define as the moment of *instrumentation*.

The study of instrumentation focuses on policy-makers' discourses, interests and rationales when selecting new instruments, as well as on the range of economic, political and institutional contingencies that condition instrument selection (Capano & Lippi, 2017; Maroy et al., 2017, Peters, 2002). Economic factors, such as the level of economic development of a country, the international economic agreements that the country in question has signed, or periods of financial crisis, encourage or inhibit the selection of certain policy instruments and determine the financial feasibility of instrument options (Lenschow, Liefferink, & Veenman, 2005). On their part, political factors such as party politics and political ideologies are also expected to frame instruments' choice. Here, Le Galès (2010) refers to Bourdieu's metaphor on the right and the left sides of the State as a way to distinguish between 'a left democratic version [of governance] promoting negotiation, and more deliberative making of the general interest and a right mode of governance using indicators, standards and technical instruments to centralize and promote a more market-oriented society' (p. 143).

Politico-administrative regimes are also core mediating factors in instrumentation processes. Considering that data-intensive policy instruments aim at altering the way public services are administered and governed, prevailing administrative institutions will expectedly condition how these instruments are received, selected and enacted. Pollitt and colleagues (2007) have studied the path-dependent reception and evolution of New Public Management (NPM) instruments in different OECD territories from this perspective (see also Gunter et al., 2016). These authors start by noticing that, after decades of NPM promotion, 'OECD public administrations have become more efficient, more transparent and customer oriented, more flexible and more focused on performance' (OECD, 2005, p. 10 in Pollitt, 2007, p. 11). Nonetheless, they observe that NPM principles (such as decentralization, outcomes-based management, accountability and competition) have crystallized quite differently according to countries' public administration traditions. Thus, in countries with a more liberal organization of the State (such as the one prevailing in most Anglo-Saxon countries), NPM has adopted a more market-oriented form. In these

countries, NPM has encouraged private sector participation in public services and more intense forms of competition between providers.

In contrast, in countries with a neo-Weberian State, which is the administrative model that mostly prevails in continental and northern Europe, NPM reforms have contributed to make services more citizen- and results-oriented, but not via the drastic promotion of market competition and choice. In these countries, the State adopts a proactive role as a facilitator of solutions to social problems and is eager to preserve the ideas of civil service and professionalism in public services (Pollitt, 2007).

Finally, within the so-called Napoleonic administrative tradition, which prevails mainly in Southern European countries and is characterized by centralized, hierarchical and uniform bureaucracies, NPM reforms 'have been tried, but with disappointing results' (Pollitt, 2007, p. 12). Components of NPM have been adopted disconnectedly and implemented unevenly. Furthermore, NPM reforms have met serious opposition from different flanks, which go from street-level to high-level bureaucrats (Kickert, 2007).

On the evolutionary dimension of policy instruments

The evolution and future use(s) of policy instruments are conditioned by the previous instruments in place. Policy change often happens through the sedimentation of policy instruments. It operates as a *layering* process, in which policy change entails the addition of new instruments on top of existing ones; but it might also act as a *conversion* process in which instruments are adapted to new circumstances over time, and new uses and purposes are given to them (Thelen, 2004; Vetterlein & Moschella, 2014).

Policy instruments might change or evolve by inertia (rather than by design), reasons as to why their evolution and effects are often unpredictable (Mahoney, 2000). Instruments 'have impacts on their own, independent from the policy goals' (Le Galès, 2010, p. 151), or from 'the decisions that created them' (Kassim & Le Galès, 2010, p. 11). As a consequence, it is difficult to predict the form that any instrument will end up assuming, as well as their most direct effects (Bezes, 2007). To a great extent, actors' responses to new instruments might be more creative and diverging than policy-makers expect and, accordingly, both the responses and the effects produced by policy instruments 'depend on how the aims and purposes ascribed to them, and the meanings and representations

they carry, are perceived, understood and responded to by key actors' (Skedsmo, 2011, p. 7).

Another important premise to understand the (unpredictable) evolution of policy instruments is that, once selected, instruments privilege certain actors and their interests (over others), and usually incentivize the generation of new constituencies. New constituencies are comprised of political and/or economic actors 'oriented towards developing, maintaining and expanding a specific instrumental model of governing' (Simons & Voß, 2018. p. 31) that have vested interests in the intensification of instrument uses. At the same time, however, new instruments might disserve other groups and, accordingly, trigger critical reactions and different forms of resistance.

To conclude, the perspective to policy instruments analysis presented here involves the systematic study of the sequence of contingencies (institutional, political, economic), events, and actions behind policy instruments' choice and their changing uses, and enquires into how and to what extent the selection and evolution of instruments contribute to advance substantive changes in the governance of public policy systems.

Methods

The paper draws on the results of a systematic literature review of indexed publications focusing on the political economy of educational governance reforms. A systematic literature review (SLR) is a methodology oriented toward the synthesis of existing research on a particular topic and is characterized by the use of explicit and transparent methods of search and selection in order to reduce possible bias (Petticrew & Roberts, 2006).

Our research followed the conventional steps of SLRs as established by specialized literature. The main source considered was the SCOPUS database, although we also relied on recommendations from key informants with expertise on countries under-represented in the indexed literature.¹ After different screening processes, the final selection of papers considered in this review includes a total of 158 papers.

Information was systematized on the basis of country-specific extraction sheets. The reason to use country-specific (instead of paper-based) extraction sheets was that, frequently, reviewed papers deal with very specific components of the assessments and

¹ The search terms can be consulted in the Appendix.

accountability framework of a country or tend to focus on policy changes in different periods of time. Aggregating fragmented pieces of literature into specific country-forms contributed to have a more comprehensive and accurate understanding of this complex education reform phenomenon.

Instruments' choice: three rationales behind educational governance reforms

The reviewed literature allowed us to differentiate between three main policy trajectories in reforming educational governance, which echo the politico-administrative traditions developed by Pollitt (2007) and Pollitt and Bouckaert (2011) sketched above.

NPM marketizers: adopting TBA to expand market competition and choice

In Anglo-Saxon countries, public sector reforms have been permeated by free-market policy principles and public choice theory. In the context of the global economic crisis of the 1970s, neoliberal ideas gained prominence in political and public policy agendas of countries such as the United States, England, New Zealand, but also Chile (see Falabella, 2015; Hursh, 2005). According to Pollit and Bouckaert (2011), these countries can be classified as 'core NPM states' or 'NPM marketizers' (p. 86), in the sense that the NPM toolkit was used strategically to advance the marketisation and the privatization of public services. In this context, education reforms adopted a market-driven approach and entailed the adoption of new governance instruments to steer 'at a distance' an increasingly complex and fragmented pool of educational providers.

The early adopters of NPM in education articulate a coherent and explicit theory of change on how TBA, parental choice and competition can trigger a sort of school improvement 'virtuous circle'. In these countries, national assessments and accountability measures have been enacted in combination with exogenous privatization policies (such as vouchers or other forms of public subsidies for private schools) with the purpose of stimulating market competition in education and empowering parents in their role as clients (Ball, 2008; Clarke, Gewirtz, & McLaughlin, 2000). NLSAs have been conceived as a pivotal policy instrument to collect data on schools' performance, to inform parental choice and to promote market accountability dynamics. Education reforms have been justified by a

persistent discourse on public schooling failures and low-quality education in the public sector, usually attributed to burdensome bureaucratic rules and absence of incentives (Falabella, 2015; Hursh, 2005; Whitty & Wisby, 2016). The necessity of stricter surveillance mechanisms has been justified by a hostile discourse against teachers and teachers' unions. Thus, in a context of mistrust of public education, TBA and standards appear as a suitable policy solution to increase State control over public schools, teachers' work and the curriculum (Whetton, 2009).

In England and Chile, these changes in educational governance have enjoyed great stability over time, in great part, due to the fact that they are part of a profound and structural process of re-structuration of the state. In fact, once enacted, TBA systems have become increasingly complex and sophisticated, and their uses have been incrementally intensified by different governments – regardless of their political ideology (Ball, 2008; Parcerisa & Falabella, 2017). In contrast, in the US, market reforms were tried in the 1980s but did not advance as quickly due to the complexities of the American political architecture (Klitgaard, 2007). In this country, school choice and TBA were not regulated as complementary policy instruments, at least at the Federal level, until the enactment of No Child Left Behind Act, passed in the year 2001 (Betebenner, Howe, & Foster, 2005). In the political process that led to the approval of the NCLB Act, discourses on competitiveness and choice were strategically combined with discourses about racial (and socioeconomic) equity and the reduction of achievement gaps (Hursh, 2005). The accountability pressures that came with NCLB opened the door to the conversion of underperforming public schools into charter schools, and widened up school exit and choice opportunities among families.

Neo-Weberian states: governance instruments travel to continental Europe

In the last two decades, governance instruments have been widely adopted in continental Europe, including central and Nordic countries where a Neo-Weberian politico-administrative regime prevails. In Neo-Weberian states, external evaluations and new accountability instruments in education were not initially chosen to promote market competition, but as a way for the central State to guarantee quality standards in a context of highly decentralized education systems. In the 1980s and 1990s, Nordic countries like Denmark, Sweden or Norway went through profound decentralization processes which transferred numerous competences and responsibilities on education to local governments

(Hatch, 2013; Moos, 2014). Decentralization in public services was adopted for subsidiarity reasons and as a way to make services more responsive to citizens' demands. However, decentralization came at the cost of central control and key stakeholders started questioning whether the national government had the necessary tools in place for monitoring the quality of education (Tveit, 2009).

Nonetheless, countries like Norway and Denmark did not react to quality control concerns until the first Programme for International Student Assessment (PISA) results were released in 2001. Disappointing results in PISA reinforced the social perception of a 'learning crisis' and opened an important window of opportunity for education reform advocates. As a consequence, most Nordic countries embraced an outcomes-based management approach to education and introduced more centralized (and standards-oriented) curricula (Møller & Skedsmo, 2013; Ydesen, 2013). They also adopted external evaluations as a way to both regain control over the curriculum and monitor the educational work of both local governments and schools more closely. Quality assurance schemes, national systems for school inspection and new education assessment units were also created (Helgoy & Homme, 2007).

In Central Europe, countries like Germany, Austria and The Netherlands have gone through a similar reform process, and have introduced standardized evaluation and accountability systems also as a way to regain control over curriculum delivery and academic results. Germany is one of the countries in Europe that has been more profoundly affected by unexpected low results in international large-scale assessments. In this country, the first PISA report generated a 'shock' that entailed the introduction of an output-oriented governance approach, which crystallized in three main interventions, namely the establishment of centralized monitoring of education standards, the strengthening of school autonomy, and the expansion of research-based policy-making. These reforms were more intense in those German *länder* (or states) at the bottom of the inner-German PISA ranking (Niemann, Martens, & Teltemann, 2017). Nonetheless, despite the emphasis on accountability in education in the aftermath of PISA, the accountability systems adopted in Germany, as happened in most Neo-Weberian states, were predominantly low-stakes (Thiel, Schweizer, & Bellmann, 2017).

In a nutshell, in Central and Northern Europe, the adoption of new assessment and accountability practices is largely motivated by a sense of an international race for educational results. In many ways, in these countries, as stated by Browes and Altinyelken (2018) in relation to the Dutch case, accountability reforms ‘can be understood as a rebalancing of the system, a “catch-up response” to the decentralization reforms that preceded them’ (p. 13).

Napoleonic states: the partial and uneven adoption of governance reforms

In Napoleonic states, which is the public administration model that prevails in Southern Europe, reforms to modernize public administration have been repeatedly tried, but have not always generated the expected changes (Gallego, 2003; Kickert, 2007). In these states, managerial education reforms have been adopted much later than in other education settings, without sufficient political backing, and not always following a comprehensive reform plan. In addition, once adopted, the implementation of accountability instruments has been uneven and highly conditioned by political contestation and economic junctures. In Napoleonic states, most teachers have civil servant status and enjoy, *de facto*, high levels of autonomy. External evaluations have usually been adopted as a way to address an administrative sense of lack of control, and to encourage school improvement dynamics (Carvalho & Costa, 2017). To a great extent, new governance instruments and techniques have been adopted as a way to modernize the governance of the education system and to adhere to international norms and discourses on educational reform. However, these changes in the governance of education are not adopted as part of a cohesive and openly deliberated reform package, and their implementation has been often discontinued (Serpieri, Grimaldi, & Vatrella, 2015; Stamelos, Vassilopoulos, & Bartzakli, 2012). As we develop below, in Italy and in some Spanish regions, programs of merit-based pay, teacher/principal evaluation and school rankings have been repeatedly piloted, but discontinued after first attempts.

In the political discourse that predominates in the South-European region, new policy instruments, techniques and tools are usually attached to a rhetoric of ‘quality assurance’, but the *theory of change* of how these instruments are expected to generate quality gains is not always well articulated and explicit. Another characteristic of educational reforms in this region is the strong emphasis on ‘school autonomy’, which is usually translated into the promotion of a more hierarchical leadership style in schools (Dobbins & Christ, 2017). The professionalization and empowerment of the school principal figure represents a

significant shift in countries with a long legacy of democratic and horizontal educational governance (with the principal being a *primus inter pares*, and many school decisions relying on community/families' participation) (Gairín Sallán, 2015). Democratic school governance (including the direct participation of families and teachers in core school decisions) in countries such as Portugal, Greece and Spain emerged as a reaction to decades of authoritarian regimes, which closely controlled the educational system for political reasons. Nonetheless, more recently, education reform advocates portray democratic school governance as ineffective and promote the introduction of managerial changes and new leadership styles in schools (Veloso, Abrantes, & Craveiro, 2013; Verger & Curran, 2014).

New and changing uses: the evolution of NLSAs and TBA systems

Besides the logic of instruments' choice, inquiring into the unfolding and evolutionary dimension of policy instruments is also necessary to understand the impact of instruments on educational governance. This section analyses the evolution of NLSAs and TBA systems, separately due to the fact that, from the perspective of their uses, NLSAs and TBA represent two analytically distinct varieties of instruments: respectively, instruments aimed at collecting information and instruments aimed at shaping behavior (Hood, 2007). Consequently, we should expect that both the pace and the nature of their evolution differ.

NLSAs: an ever-expanding instrument?

Data collection instruments such as NLSAs tend to remain relatively stable over time and are rarely reversed. The main change experienced by NLSAs is related to their sophistication and expansion. With the passage of time, the frequency and scope of national assessments tend to increase. Changes in assessments are frequently the result of concerns on the reliability or relevance of the tests – for instance, when changes are introduced in order to ensure better alignment with curricular standards, or when the evaluated subjects are expanded over concerns about the narrowing of the curriculum. These preoccupations are often triggered by the increasing number of purposes given to (or expected to be served by) large-scale assessments, turning test validity into a matter of critical importance.

This intensification pattern holds across different contexts, but is particularly evident in core NPM countries, such as Chile and England. In Chile, during the late 1990s and early 2000s, as concerns over education quality became central to the education policy agenda, a series of changes were implemented to align national curriculum and assessment (Gysling, 2015). The tests were progressively reoriented towards the evaluation of a national curricular framework, in which cognitive skills of a superior order were measured through the introduction of new types of questions. Between 2005 and 2011, new grades and subjects were added to the national testing framework, and the test frequency was intensified so to allow student tracking over time (Bravo, 2011).

In other cases, the intensity of the testing framework is not necessarily altered, but new monitoring tools are devised – for instance, by adding new levels of data (dis)aggregation, creating new schools-classification systems, or developing new composite measures (indexes, typologies, etc.). The introduction of additional measures and more sophisticated tools is the logical consequence of the creation of specialized organizations (typically more independent evaluation agencies) whose main responsibility is to supervise and make use of a growing volume of collected data. These dynamics can be clearly observed in the English testing framework. While national assessments have remained relatively stable (in terms of number and frequency) since their introduction with the 1988 Education Reform Act (ERA), the number and sophistication of performance-related information has increased substantially – for instance with the adoption of the Pupil Achievement Trackers and Performance and Assessment (Panda) reports. In addition, in 2010, a series of decisions contributed to the consolidation and expansion of the national testing framework. This included the introduction of a new test (Phonics Screening Check, age 6), the revision of the Early Years Foundation Stage Profile (age 5), and the establishment of a new performance measure (English Baccalaureate, which highlights the proportion of pupils of a given secondary school achieving high grades in the General Certificate of Secondary Education) (Bradbury, 2014; Mansell, 2011).

The US provides another example of the expansive nature of national assessments. The passing of NCLB in 2001 established an extensive testing framework by requiring states to test students in grades 3 through 8, and once in high school. Since then a number of techniques that tie student performance to teacher evaluations, as incentivized by the Race to the Top Act of 2011, have also emerged. Among them, value-added models feature

prominently given its widespread use by several states and urban districts (Amrein-Beardsley & Holloway, 2017; Baker, Oluwole, & Green, 2013).

The ebb and flow of test-based accountability

TBA systems experience a slower development than NLSAs, as they unfold gradually, and are more likely to undergo a rather uneven evolution, with some of their components being discarded after some time in place. Nonetheless, these dynamics play out quite differently according to the different administrative traditions sketched above.

In the case of *NPM marketizers*, stakes have primarily tended to increase and intensify. This occurs not only because reputational and market consequences tend to be reinforced over time, but also because administrative and bureaucratic stakes are likely to be added to the accountability system. In Chile, for instance, the national test was initially created with the aim of informing parents' choice and for curricular control purposes (Meckes & Carrasco, 2010, Falabella, 2015). However, soon after the first publication of test results in 1995, a salary bonus linked to schools' performance and a series of additional administrative sanctions and dispositions were adopted (Flórez, 2013). In the 2000s, school subsidies for low-income students became conditioned to the school compliance with State-defined learning goals, and low-ranked schools became more closely supervised, lost the capacity to autonomously administer public funding, and risk closure if they did not show signs of improvement. Remarkably, some of these new administrative dispositions were adopted by the center-left government as a means to correct market failures and reinforce the role of the State as a regulatory agent. At the same time, these same dispositions were supported by the Right because they perceived them as a way to reform education, but preserve and secure the market system at the same time (Parcerisa & Falabella, 2017).

Similarly, in England, different authors document a rise in the stakes associated to standardized tests during the late 1990s and early 2000s, under the New Labour government. Some of the consequences of the evaluations had a reputational nature – since they included the public posting of performance data – and were paralleled by the intensification of the school choice and competition agenda (Mansell, 2011; Muijs & Chapman, 2009). However, other consequences of a more material and administrative nature were added as well. New arrangements included 'light-touch' intervention policies for failing or low-performing schools as well as performance-based pay schemes (Mansell,

2011; Whetton, 2009). These changes entailed a major departure from the accountability system put in place with the 1988 ERA, which relied essentially on market dynamics. However, a reform passed by the Coalition government formed by Conservatives and Liberal Democrats after the 2010 election meant a certain return to a market-based TBA system. New dispositions reduced the role and intervention capacity of the inspectorate in well-functioning schools, and those scoring as outstanding were deemed exempt from regular inspection (Mansell, 2011).

In so-called *Neo-Weberian states*, an intensification pattern of TBA measures is less clearly discernible. In fact, in some of these cases, it is even possible to detect a certain deceleration of the accountability agenda, resulting from the removal of the initially adopted market and reputational consequences. Although many Nordic countries have experimented with the publication of test results, most of them have finally stopped doing so, at least at the national level, due to concerns with the quality and reliability of the measures, but also because of the critical response of key stakeholders.

The case of Norway is illustrative of these logics of adjustment and reversal. In 2005, a Conservative-led coalition introduced a National Evaluation System that included for the first time a national standardized test. The Conservative government behind the reform expected to create pressure on schools through the combination of the public posting of test scores and the introduction of greater levels of school choice. Public posting of school results was however suspended in 2005, once the new center-left coalition came into power (Camphuijsen, Skedsmo, & Møller, 2018; Hatch, 2013). A similar pattern has been documented in Denmark. In 2005, a public order required schools to publish relevant information on their websites, including results from evaluation and teaching (also, the 2002 Act on Transparency had established similar requirements), and the Ministry of Education experimented with the publication of school rankings on its website. However, the publication of school results remained a particularly controversial issue, and the center-left coalition that came to power in 2011 suspended the publication of school league tables (Ydesen & Andreasen, 2014).

However, in both Denmark and Norway, the publication of school results re-emerged, even in the absence of national governmental action. In these two countries, even when the national government does not publish school scores, many local governments or the

media do so, taking advantage of transparency rules in public administration. Similarly, in British Columbia (Canada), it is the Fraser Institute – an advocacy-oriented think tank explicitly committed to a deregulation and marketization agenda – that produces school report cards disseminated through the media (Simmonds & Webb, 2013).

In *Napoleonic States*, TBA reforms appear to advance only through a trial-and-error logic and a back-and-forth dynamic characterized by frequent discontinuities. Test-based accountability remains relatively underdeveloped, and the raising of the stakes depends largely on the political orientation of the government in power. Attempts to introduce some form of principal evaluation in Italy illustrate these dynamics well. During the early 2000s, a series of principal-evaluation pilots were developed by a center-right government. Although evaluation was voluntary, the initiative faced high rates of rejection – partially as some assumed that the pilots were only the first step before the implementation of mandatory evaluation. The initiative was discontinued in 2006, as a center-left coalition took office. However, the election of a new center-right government in 2008, along with the appointment of a Minister of Education explicitly committed to meritocracy in education, put the principal evaluation back to the policy agenda (Grimaldi & Serpieri, 2013). During the early 2010s, a new project establishing the voluntary evaluation of schools and teachers was proposed, giving rise in 2012 to an experimental evaluation model integrating the measurement of the school added value and the evaluation of principals (Serpieri et al., 2015). It was not until the approval of the education reform *La Buona Scuola*, passed in 2015 by a technocratic cabinet, that principal evaluation was consolidated as part of the new accountability framework (Montefiore, 2018).

Similar dynamics can be observed in Madrid, an autonomous region of Spain that, during the early 2000s introduced a series of market-based accountability measures, including the introduction of a standardized and census-based test combined with an increase in school choice freedom. With these policies in place, this region took an ‘outlier’ market trajectory that, apparently, did not fit well with the main characteristics of a Napoleonic administrative tradition. The adoption of these reforms owed much to the strong leadership, entrepreneurship and top-down government style of President Aguirre (2003–2012), who was strongly and personally committed to market freedoms in all types of sectors, including education. However, the reform was not resilient to Aguirre’s resignation in 2012, and the new government – even when it was in hands of the same political party

– abandoned some of the most emblematic market-accountability dispositions, including the publication of school results, and reduced the frequency of testing (Pagès & Prieto, 2018).

Finally, federal or highly decentralized states deserve separate consideration. In these cases, the progressive heightening of the stakes generally occurs at the sub-national level. Hence, even if the Federal government does not associate material or bureaucratic consequences to national assessments, local or state authorities can take advantage of these instruments being in place in order to adopt their own accountability measures (see for instance Gable & Lingard, 2015; Termes & Mentini, 2018).

New constituencies: the emergence of economic and political subjects within educational governance reforms

Economic interests in testing

Governance reforms in education have contributed to the emergence of a testing and measurement industry. The presence of this industry is bigger in those countries defined as NPM marketizers, where testing is more intensive and is usually attached to higher-stakes. In these countries, private companies, consultancies and research organizations such as Pearson, the Australian Council for Educational Research (ACER) or the Learning Bar benefit from substantive contracts with governments for the design, administration and/or data analysis of national assessments (Burch, 2006). These and other companies also sell school improvement services, lesson plans and/or educational platforms to those local governments and schools that aim at strengthening their performance (Hogan, Sellar, & Lingard, 2016). School improvement companies are very active in countries such as England and Chile, and many specialize in how to increase students' scores in external evaluations. School improvement is an important market niche in Chile, to a great extent, because schools can resort to public funding to hire these types of services (Osses, Bellei, & Valenzuela, 2015).

The expansion of economic interests in testing and measurement activities is key to understand the on-going spread of external evaluations and related accountability instruments. As the OECD (2013) acknowledges, the fact that 'standardized student assessment becomes a more profitable industry' means that 'companies have strong

incentives to lobby for the expansion of student standardized assessment as an education policy, therefore, influencing the activities within the evaluation and assessment framework' (p. 51).

Nonetheless, public universities and research institutes are also involved in contracts for test design and data analysis, but their presence is relatively bigger in Continental Europe than in the Anglo-Saxon world. In many European countries, public agencies are centrally involved in testing-related activities. In England, universities and academic centers are also involved in test development, but smaller academic suppliers have been largely withdrawing from testing arrangements due to the inclusion of stricter conditions in the contracts, such as penalties for not meeting the deadlines (Whetton, 2009).

Overall, the increasing involvement of both public and increasingly private groups within education testing activities (both as third-party producers and as parties supporting policy implementation) explains why testing and TBA instruments expand not only territorially, but also toward new areas of educational activity and education levels. When the political and economic interests of these groups are strong, there are more reasons to expect that these policy instruments will endure in time, independently of their effectiveness (see Dale, 2018). In other words, the political and economic positioning of the groups that administer governance instruments – or deliver related services – within the education policy field is key to explain policy continuity and lock-in effects, even when the effectiveness of intensive uses of standardized testing is increasingly questioned by academic evidence.

Spreading like wildfire, but meeting firewalls

Standards, assessments and accountability are known for 'spreading like wildfire', among other reasons, because they are inexpensive to fund in comparison to other policy alternatives (Hargreaves & Shirley, 2009). The fact that the adoption of policy instruments such as NLSAs or TBA results from debates of a technical (rather than political) nature also makes the articulation of social responses difficult. National assessments tend to generate less resistance because they are usually adopted as non-intrusive data-gathering instruments, and their ultimate effect on teachers and schools is not evident since the outset. However, under certain circumstances, these instruments also generate critical reactions.

TBA is the most contentious governance instrument, especially when associated with high-stakes outcomes. High-stakes accountability usually triggers passionate debates between its supporters and detractors, and is also a motive of collective action (Pizmony-Levy & Woolsey, 2017). Nonetheless, responses to TBA are not always led by teachers' unions (TUs). In fact, in the light of existing literature, TUs responses to accountability reforms cannot be taken for granted. On the contrary, accountability usually places TUs in difficult political dilemmas, a major reason why their responses are variegated and not always as defiant as could be expected. For example, in Chile the main TU (*Colegio de Profesores*) did consent – and participated in the definition of – new forms of teacher evaluation policies. The union saw these policies as a lesser evil in a context of profound marketization and pauperization of teachers' work, and as a *de facto* opportunity to raise teachers' salaries (Gindin & Finger, 2013; Vaillant, 2005).

In the context of the US, the main unions were very cautious regarding TBA under the NCLB Act and, despite their general dissatisfaction with the reform, they could not agree on a unitary response (Hursh, 2005). Furthermore, given the fact that NCLB had such a strong equity discursive frame, the leadership of the unions wanted to avoid being seen as insensitive to children learning issues or to existing learning gaps in front of society. The teachers' critique to standardized testing is particularly challenging in the US because, first, the societal trust in teachers is relatively low and, second, standardized testing is sound with values that are deeply rooted in American society such as meritocracy, achievement and effort (Au, 2016).

In Nordic European countries, TUs opposition to TBA is not only explained by ideological reasons, but by normative and professional understandings of how the teaching profession should be regulated. In Norway, teachers adopted a critical attitude towards NLSAs policies because they felt their professional autonomy and judgment capacity was being challenged, although they did not articulate a confrontational campaign against the national assessment first implemented in 2004. Their critical position against this assessment was not only ideologically motivated, but emphasized the poor design and quality of the test (Iveit, 2009). In contrast, in Sweden, NLSAs did not generate so much controversy. There, teachers did not see the national assessment as an external instrument challenging their autonomy, but as a useful tool to support their professional development and to establish quality standards in education (Helgoy & Homme, 2007).

A more oppositional approach among TUs appears in Southern Europe. In Italy, the strong bargaining capacity of the TUs is crucial to understand the discontinuities and changes experienced by many managerial programs and tools (Barzanò & Grimaldi, 2013; Grimaldi & Serpieri, 2013). In the Portuguese context, national evaluations were perceived as repressive instruments with reminiscences to the dictatorial period, and generated fierce resistance by well-organized TUs (Veloso et al., 2013). In general, in Southern Europe, TUs, but also other local stakeholders and parent associations, are especially belligerent against accountability instruments that more directly challenge teachers' autonomy and the democratic governance of schools.

The adoption of TBA also generates critical responses from social actors other than TUs, including students' movements, critical scholars, pedagogical associations and, especially, parents. In fact, these social responses tend to emerge in contexts where TUs adopt an ambiguous or passive position in front of standardized testing and accountability reforms. For example, in Norway, the government approved a moratorium in standardized testing after students' boycotts were carried out in the main cities of the country. This boycott was supported by teachers, but only covertly (Helgoy & Homme, 2007).

In the US, the rapid intensification of standardized testing has triggered the emergence of an ideologically transversal opt-out movement that crystallizes in organizations such as *FairTest* or *United Opt Out* (Dobrick, 2014; Pizmony-Levy & Woolsey, 2017). In states like New York, families have massively followed calls to boycott the test and, in response to these actions, the government has introduced improvements in the test design and administration. The political influence of this movement also manifests in the fact that 'opt-out parents' have been elected in different district school boards (Wang, 2017).

More recently, similar social movements that boycott national tests have emerged in other countries, such as Spain (see Saura et al., 2017), and Chile (Campos-Martínez & Guerrero-Morales, 2016; Pino-Yancovic, Oyarzún-Vargas, & Salinas-Barrios, 2016). These social movements, which tend to be led by middle-class families, articulate a very sophisticated narrative regarding the non-desired educational effects of standardized testing, and are particularly skillful in managing social media and performing innovative collective actions.

Discussion and conclusions

In the last decades, OECD and middle-income countries have given greater salience to external evaluations, targets, standards and accountability in the governance of their educational systems. Policy communities share, internationally, a similar discourse on evaluation and accountability and, at the regulatory level, we have observed that countries are almost unanimously adopting NLSAs and TBA instruments when reforming their educational systems. However, both the instrumentation process (i.e., the political process through which policy instruments are being adopted) and the uses that governments are giving to policy instruments are not converging so clearly cross-nationally.

Institutional legacies strategically mediate the adoption of education governance instruments. It is mainly in *Anglo-Saxon countries*, with a liberal organization of the State, where these instruments have been adopted with a more obvious pro-market purpose. In these countries, evaluation and accountability instruments are explicitly used to promote school competition and choice, and are more clearly attached to school rankings and merit-based pay formulas. With the passage of time, accountability systems have become more complex and sophisticated, and their stakes higher. The more established political forces (Liberal, Conservative and Labour parties) agree on the central role of testing and TBA in educational reforms, and on the main uses that should be given to these instruments. Accordingly, the public posting of schools' results is less likely to be questioned and, in fact, the production of school rankings has frequently been promoted by center-left parties, in an attempt to make public services more transparent and democratize school choice. A testing industry has emerged more strongly in the context of NPM marketizers, and TUs responses to these policies has been rather timid.

In contrast, in *neo-Weberian states*, evaluation and accountability instruments have been adopted following a quality assurance rationale, in an education policy landscape characterized by high levels of decentralization. The main political parties agree on the adoption of assessments and accountability systems, but not always on their uses, with the right more inclined to produce rankings and promote market competition than the left.²

² Although the reviewed literature also shows that social-democratic governments tend to be more belligerent with certain TBA uses when in opposition than when in power (see for instance Solhaug, 2011).

Accordingly, the uses of governance instruments vary in different legislative terms, but are also highly contingent to local politics.

Finally, in *Napoleonic states*, accountability has been adopted at the regulatory level and to comply with international norms and discourses on educational governance, but its enactment is very uneven and frequently remains in a latent or incipient stage. In these countries, the advancing of governance reforms has been more clearly conditioned by political and economic junctures. The political consensus (between the right and the left) around external assessments and accountability is not as evident as in the previous cases, and TUs have confronted more directly TBA and other managerial reforms, usually through industrial action. Overall, in both Neo-Weberian and Southern European countries, governance instruments are not always adopted as a synonym of a market competition agenda, but tend to generate increasing performative pressures and unrest among key education stakeholders.

The main political forces advocating the adoption of national assessments and accountability instruments also differ across contexts. According to the literature reviewed, in mainland Europe, the OECD is at the center of a transnational *instrument constituency* (cf. Béland & Howlett, 2016) that effectively advocates school autonomy with accountability as a desirable form of educational governance. There, the OECD, mainly via PISA, but also through other policy mechanisms, has strongly triggered national educational debates that have derived into significant changes in the education sector. In contrast, in core NPM states, the origin of educational governance reforms is more endogenous and, accordingly, the role of domestic policy entrepreneurs, consultants and think tanks is more often reported than the role of international organizations.

As a note of caution, the policy trajectories that we have identified are not exhaustive. They do not reflect all possible reform manifestations, but those that are more widely represented in existing literature. As more research on the topic is being produced in different world regions, the more feasible it will be to complement and widen our categories. Furthermore, although the policy trajectories are informed by specific public administration regimes, these same administrative regimes cannot be taken as a 'kind of unchanging bedrock' (Pollitt & Bouckaert, 2011, p. 48, cited in Gunter et al., 2016, p. 16).

Federal countries are particularly difficult to classify from the administrative regimes' perspective, because their own states might lead to diverging reform trajectories.

On the cumulative nature of educational governance reforms

The educational transformations that we have analyzed in this paper have a fragmented and cumulative nature in the sense that, in most cases, policy change advances as the result of the sedimentation and layering of different instruments, techniques and tools that are not necessarily articulated in a predefined reform program. With the passage of time, instruments such as accountability systems gain autonomy from their promoters and, accordingly, adopt functions and generate effects that were not initially foreseen.

Furthermore, with the exception of NPM marketizers such as Chile and England, NLSAs and TBA are generally adopted in different points in time and, in fact, as the result of differentiated debates, or in response to different demands. The implementation of a NLSA system generally precedes the adoption of TBA measures (since some form of census-based assessment is a necessary condition for TBA to happen) and, with the passage of time, economic, bureaucratic or reputational consequences tend to be attached to the assessments. Test consequences are usually State-mandated, but many other consequences emerge independently from governmental policies and intentions (e.g., the role of the media or philanthropic organizations in the dissemination of test scores). Overall, the cumulative nature of governance reforms makes it particularly necessary to get an accurate, sequential and diachronic understanding of the development of this model of educational change.

Our results also suggest that policy instrumentation is not a fixed moment in time, and goes through recurrent back-and-forth dynamics. These dynamics are, on occasions, the result of changing political and economic junctures, but also the result of effective collective actions (such as those organized by TUs or the opt-out movement). Nonetheless, even if collective action might alter the uses (or the design) of policy instruments, it is less likely to challenge the very existence of these instruments.

The layering process through which national assessments and TBA evolve explains why it is so difficult to challenge the presence of these instruments in educational systems. Key stakeholders do not foresee many of the changes that these instruments involve from the

beginning, and they only start reacting to them once their uses have intensified and been routinized. Furthermore, our findings reflect that there are several factors that have a lock-in effect on the use of external evaluations and accountability mechanisms. These include the emergence of new constituencies and economic interests around testing activities; the international education environment that, among other things, favors an international education race for constantly-improving learning results; and the broad political consensus generated around data-intensive governance instruments.

The malleable nature of governance instruments (i.e. they might serve to address learning gaps and strengthen the democratic control of education, but also to promote school choice and competition) contributes to their widespread and politically transversal adoption, as well as to their incremental evolution. From the perspective of instrumentation, external evaluations and accountability are appealing and convenient choices. The seductive power of these instruments relies on the fact that they contribute to transform complex and multi-dimensional educational realities into numerical categories, and to construct the perception that deep educational problems (such as inequalities or quality issues) can be addressed by setting up predefined patterns of conduct, measuring actors' performance, and distributing incentives accordingly (Falabella, 2018, Barbana, Dupriez, Dumay, 2014).

For all these reasons, data-intensive policy instruments continue expanding globally and are gaining centrality in the governance of education systems. This globalizing phenomenon urges us to retrieve new sources of evidence on the intended and unintended effects of educational governance reforms in schools' micro-dynamics. Future research on the topic could explore the multiple appropriations of governance instruments at the school level, and analyze under what conditions these instruments generate more instrumental or expressive responses among teachers and principals. This type of research, which is particularly underdeveloped in relation to soft-accountability systems, could contribute to inform better policy decisions in the future. By unraveling the political connotations and implications of policy instruments, research can also contribute to promote more democratic and informed debates about educational change in the governance era.

Disclosure statement

No potential conflict of interest was reported by the authors

References

- Amrein-Beardsley, A., & Holloway, J. (2017). Value-added models for teacher evaluation and accountability: Commonsense assumptions. *Educational Policy*. <http://doi.org/10.1177/0895904817719519>
- Au, W. (2016). Meritocracy 2.0: High-stakes, standardized testing as a racial project of neoliberal multiculturalism. *Educational Policy*, 30(1), 39–62.
- Baker, B. D., Oluwole, J. O., & Green, P. C. (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the Race-to-the-Top era. *Education Policy Analysis Archives*, 21(5). <http://doi.org/10.14507/epaa.v21n5.2013>
- Ball, S. J. (2008). The legacy of ERA, privatization and the policy ratchet. *Educational Management Administration & Leadership*, 36(2), 185–199. <http://doi.org/10.1177/1741143207087772>
- Ball, S. J., Junemann, C., & Santori, D. (2017). *Edu.net. Globalisation and education policy mobility*. London: Routledge. <http://doi.org/10.4324/9781315630717>
- Barbana, S., Dumay, X., & Dupriez, V. (2014). Perceptions et usages des instruments d'accountability. Enquête exploratoire dans l'enseignement secondaire en Belgique francophone. [Perceptions and uses of accountability instruments. Exploratory survey in secondary education in French-speaking Belgium.] *Éducation compare*, 12, 21-44.
- Barzanò, G., & Grimaldi, E. (2013). Discourses of merit. The hot potato of teacher evaluation in Italy. *Journal of Education Policy*, 28(6), 767–791. <http://doi.org/10.1080/02680939.2013.774439>
- Béland, D., & Howlett, M. (2016). How solutions chase problems: Instrument constituencies in the policy process. *Governance*, 29(3), 393–409.
- Betebenner, D. W., Howe, K. R., & Foster, S. S. (2005). On school choice and test-based accountability. *Education Policy Analysis Archives*, 13(41). <http://doi.org/10.14507/epaa.v13n41.2005>
- Bezes, P. (2007). The hidden politics of administrative reform: Cutting French civil service wages with a low-profile instrument. *Governance*, 20(1), 23–56. <http://doi.org/10.1111/j.1468-0491.2007.00343.x>

- Bradbury, A. (2014). 'Slimmed down' assessment or increased accountability? Teachers, elections and UK government assessment policy. *Oxford Review of Education*, 40(5), 610–627. <http://doi.org/10.1080/03054985.2014.963038>
- Bravo, J. (2011). SIMCE: Pasado, presente y futuro del sistema nacional de evaluación [SIMCE: Past, present and future of the national system of evaluation]. *Estudios Públicos*, 123, 189–211.
- Browes, N., & Altinyelken, H. K. (2018, September). The evolution of test-based accountability in the autonomous Dutch system: A policy instruments approach. Paper presented at the European Conference on Educational Research (ECER), Bolzano, Italy.
- Burch, P. E. (2006). The new educational privatization: Educational contracting and high stakes accountability. *Teachers College Record*, 108(12), 2582–2610.
- Campos-Martínez, J., & Guerrero-Morales, P. (2016). Efectos indeseados de la medición de la calidad educativa en Chile. La respuesta de la sociedad civil [Undesired effects of quality measuring in Chile. The civil society response]. *Cadernos CEDES*, 36(100), 355–374. <http://doi.org/10.1590/cc0101-32622016171351>
- Camphuijsen, M. Skedsmo, G., & Møller, J. (2018, September). School autonomy with accountability as a global education reform: Its adoption and re-contextualization in the Norwegian context. Paper presented at the European Conference on Educational Research (ECER), Bolzano, Italy.
- Capano, G., & Lippi, A. (2017). How policy instruments are chosen: Patterns of decision makers' choices. *Policy Sciences*, 50(2), 269–293. <http://doi.org/10.1007/s11077-016-9267-8>
- Carvalho, L. M., & Costa, E. (2017). Avaliação externa das escolas em Portugal: atores, conhecimentos, modos de regulação [External evaluation of schools in Portugal: Actors, knowledges and regulation modes]. *Revista Brasileira de Política e Administração Da Educação*, 33(3), 685–705. <http://doi.org/10.21573/vol33n32017.79302>
- Clarke, J., Gewirtz, S., & McLaughlin, E. (2000). *New managerialism, new welfare*. Thousand Oaks: SAGE Publications.
- Dale, R. (2018). Global education policy: Creating different constituencies of interest and different modes of valorisation. In A. Verger, H. K. Altinyelken, & M. Novelli (Eds.), *Global education policy and international development: New agendas, issues and policies* (pp. 289–298). New York: Bloomsbury.

- Dobbins, M., & Christ, C. (2017). Do they matter in education politics? The influence of political parties and teacher unions on school governance reforms in Spain. *Journal of Education Policy*. <http://doi.org/10.1080/02680939.2017.1406153>
- Dobrick, A. (2014). Poverty and pretense: Good intentions and misguided educational reform from No Child Left Behind through Race to the Top. In E. M. Zamani-Gallaher (Ed.), *The Obama administration and educational reform* (pp. 27–44). Bingley: Emerald Group Publishing Limited. <http://doi.org/10.1108/S1479-358X20130000010002>
- Falabella, A. (2015). El mercado escolar en Chile y el surgimiento de la nueva gestión pública: el tejido de la política entre la dictadura neoliberal y los gobiernos de la centroizquierda. [The school market in Chile and the rise of new public management: the political intersection between the neoliberal dictatorship and center-left governments.] *Educação & Sociedade*, 36(132), 699–722. <http://doi.org/10.1590/ES0101-73302015152420>
- Falabella, A. (2018). The seduction of hyper-surveillance: Standards, testing and accountability policies in Chilean schools. Paper presented at the American Educational Research Association (AERA) Annual Meeting, New York, United States, April 13th-17th.
- Flórez, M. T. (2013). *Análisis crítico de la validez del sistema de medición de la calidad de la educación (SIMCE)* [Critical analysis of the validity of the system of measurement of education quality (SIMCE)]. Oxford University Centre for Educational Assessment: Oxford.
- Gable, A., & Lingard, B. (2015). NAPLAN data: A new policy assemblage and mode of governance in Australian schooling. *Policy studies*, 37(6), 568–582. <https://doi.org/10.1080/01442872.2015.1115830>
- Gairín Sallán, J. (2015). Autonomy and school management in the Spanish context. *Educational, Cultural and Psychological Studies*, 11, 103–117. <http://doi.org/10.7358/ecps-2015-011-gair>
- Gallego, R. (2003). Public management policy making in Spain, 1982–1996: Policy entrepreneurship and (in) opportunity windows. *International Public Management Journal*, 6(3), 283–307.
- Gindin, J., & Finger, L. (2013). *Promoting education quality: The role of teachers' unions in Latin America*. (Paper commissioned for the EFA Global Monitoring Report 2013/2014, Teaching and learning: Achieving quality for all). Retrieved from scholar.harvard.edu/files/lesliefinger/files/unesco_paper.pdf

- Grimaldi, E., & Serpieri, R. (2013). Jigsawing education evaluation. Pieces from the Italian New Public Management puzzle. *Journal of Educational Administration and History*, 45(4), 306–335. <http://doi.org/10.1080/00220620.2013.822350>
- Gunter, H. M., Grimaldi, E., Hall, D., & Serpieri, R. (Eds.). (2016). *New public management and the reform of education: European lessons for policy and practice*. New York/London: Routledge.
- Gysling, J. (2015). The historical development of educational assessment in Chile: 1810–2014. *Assessment in education: Principles, Policy & Practice*, 23(1), 8–25. <http://doi.org/10.1080/0969594x.2015.1046812>
- Hall, P. A. (1993). Policy paradigms, social learning, and the State: The case of economic policymaking in Britain. *Comparative Politics*, 25(3), 275–296. <http://doi.org/10.2307/422246>
- Hargreaves, A., & Shirley, D. (2009). *The fourth way. The inspiring future of educational change*. Thousand Oaks: Corwin.
- Hatch, T. (2013). Beneath the surface of accountability: Answerability, responsibility and capacity-building in recent education reforms in Norway. *Journal of Educational Change*, 14(2), 113–138. <http://doi.org/10.1007/s10833-012-9206-1>
- Helgøy, I., & Homme, A. (2007). Towards a new professionalism in school? A comparative study of teacher autonomy in Norway and Sweden. *European Educational Research Journal*, 6(3), 232–249. <http://doi.org/10.2304/eejr.2007.6.3.232>
- Hogan, A., Sellar, S., & Lingard, B. (2016). Commercialising comparison: Pearson puts the TLC in soft capitalism. *Journal of Education Policy*, 31(3), 243–258. <http://doi.org/10.1080/02680939.2015.1112922>
- Hood, C. (2007). Intellectual obsolescence and intellectual makeovers: Reflections on the tools of government after two decades. *Governance*, 20(1), 127–144. <http://doi.org/10.1111/j.1468-0491.2007.00347.x>
- Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets and the decline in educational equality. *British Educational Research Journal*, 31(5), 605–622. <http://doi.org/10.1080/01411920500240767>
- Kassim, H., & Le Galès, P. (2010). Exploring governance in a multi-level polity: A policy instruments approach. *West European Politics*, 33(1), 1–21. <http://doi.org/10.1080/01402380903354031>

- Kauko, J., Rinne, R., & Takala, T. (Eds.). (2018). *Politics of quality in education. A comparative study of Brazil, China, and Russia*. Abingdon/New York: Routledge. <http://doi.org/10.4324/9780203712306>
- Kickert, W. (2007). Public management reforms in countries with a Napoleonic state model: France, Italy and Spain. In C. Pollitt, S. V. Thiel, V. Homburg, & S. Van Thiel (Eds.), *New public management in Europe* (pp. 26–51). London: Palgrave Macmillan UK. http://doi.org/10.1057/9780230625365_3
- Klitgaard, M. B. (2007). Do welfare state regimes determine public sector reforms? Choice reforms in American, Swedish and German schools. *Scandinavian Political Studies*, 30(4), 444–468. <http://doi.org/10.1111/j.1467-9477.2007.00188.x>
- Lascoumes, P., & Le Galès, P. (2007). Introduction: Understanding public policy through its instruments? From the nature of instruments to the sociology of public policy instrumentation. *Governance*, 20(1), 1–21. <http://doi.org/10.1111/j.1468-0491.2007.00342.x>
- Le Galès, P. (2010). Policy instruments and governance. In M. Bevir (Ed.), *Handbook of governance* (pp. 142–159). London/Thousand Oaks: Sage.
- Lenschow, A., Liefferink, D., & Veenman, S. (2005). When the birds sing. A framework for analysing domestic factors behind policy convergence. *Journal of European Public Policy*, 12(5), 797–816. <http://doi.org/10.1080/13501760500161373>
- Mahoney, J. (2000). Path dependence in historical sociology. *Theory and Society*, 29(4), 507–548.
- Mansell, W. (2011). Improving exam results, but to what end? The limitations of New Labour's control mechanism for schools: Assessment-based accountability. *Journal of Educational Administration and History*, 43(4), 291–308. <http://doi.org/10.1080/00220620.2011.606896>
- Maroy, C., Pons, X., & Dupuy, C. (2017). Vernacular globalisations: Neo-statist accountability policies in France and Quebec education. *Journal of Education Policy*, 32(1), 100–122.
- Meckes, L., & Carrasco, R. (2010). Two decades of SIMCE: An overview of the national assessment system in Chile. *Assessment in Education: Principles, Policy & Practice*, 17(2), 233–248. <http://doi.org/10.1080/09695941003696214>
- Møller, J., & Skedsmo, G. (2013). Modernising education: New public management reform in the Norwegian education system. *Journal of Educational Administration and History*, 45(4), 336–353. <https://doi.org/10.1080/00220620.2013.822353>

- Montefiore, G. (2018). The good school reform. A double analysis of the last Italian NPM Education reform in its development from government idea to law and in its rhetoric. Unpublished manuscript, Universitat Autònoma de Barcelona, Spain.
- Moos, L. (2014). Educational governance in Denmark. *Leadership and Policy in Schools*, 13(4), 424–443. <http://doi.org/10.1080/15700763.2014.945655>
- Muijs, D., & Chapman, C. (2009). Accountability for improvement: Rhetoric or reality? In C. Chapman, & H. M. Gunter (Eds.), *Radical reform. Perspectives on an era of educational change* (pp. 28–41). Abingdon/New York: Routledge.
- Niemann, D., Martens, K., & Teltemann, J. (2017). PISA and its consequences: Shaping education policies through international comparisons. *European Journal of Education*, 52(2), 175–183. <http://doi.org/10.1111/ejed.12220>
- OECD. (2005). *Modernising Government: The Way Forward*. Paris: OECD.
- OECD. (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. Paris: OECD. Retrieved from: <http://www.oecd.org/education/school/synergies-for-better-learning.htm>
- Osses, A., Bellei, C., & Valenzuela, J. P. (2015). External technical support for school improvement: Critical issues from the Chilean experience. *Journal of Educational Administration and History*, 47(3), 272–293. <http://doi.org/10.1080/00220620.2015.1038699>
- Pagès, M., & Prieto, M. (2018, September). SAWA policies in a context of school choice: Evidence from Spain. Paper presented at the European Conference on Educational Research (ECER), Bolzano, Italy.
- Parcerisa, L., & Falabella, A. (2017). La consolidación del Estado evaluador a través de políticas de rendición de cuentas: Trayectoria, producción y tensiones en el sistema educativo. [The consolidation of the evaluative state through accountability policies: Trajectory, enactment and tensions in the Chilean education system.] *Education Policy Analysis Archives*, 25(89). <http://doi.org/10.14507/epaa.25.3177>
- Peters, G. (2002). The politics of tool choice. In L. M. Salamon (Ed.), with O. V. Elliott, *The tools of government. A guide to the new governance* (pp. 552–564). Oxford: Oxford University Press.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Oxford, UK: Blackwell Publishing.
- Pino-Yancovic, M., Oyarzún-Vargas, G., & Salinas-Barrios, I. (2016). Crítica a la rendición de cuentas: narrativa de resistencia al sistema de evaluación en Chile. [A critical approach to

- accountability: a narrative of resistance against the evaluation system in Chile.] *Cadernos CEDES*, 36(100), 337–354. <http://doi.org/10.1590/cc0101-32622016171362>
- Pizmony-Levy, O., & Woolsey, A. (2017). Politics of education and teachers' support for high-stakes teacher accountability policies. *Educational Policy Analysis Archives*, 25(87). <http://doi.org/10.14507/epaa.25.2892>
- Pollitt, C. (2007). Convergence or divergence: What has been happening in Europe? In C. Pollitt, S. V. Thiel, V. Homburg, & S. Van Thiel (Eds.), *New public management in Europe* (pp. 10–25). London: Palgrave Macmillan UK. http://doi.org/10.1057/9780230625365_2
- Pollitt, C., & Bouckaert, G. (2011). *Public Management Reform. A comparative analysis-New Public Management, Governance, and the Neo-Weberian State* (3rd edition). Oxford. Oxford University Press.
- Ramirez, F. O., Schofer, E., & Meyer, J. W. (2018). International tests, national assessments, and educational development (1970–2012). *Comparative Education Review*, 62(3), 344–364. <http://doi.org/10.1086/698326>
- Rizvi, F., & Lingard, B. (2010). *Globalizing education policy*. London, UK: Routledge.
- Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In K. Mundy, A. Green, B. Lingard, & A. Verger (Eds.), *The handbook of global education policy* (pp. 128–144). West Sussex: Wiley-Blackwell.
- Saura, G., Muñoz-Moreno, J. L., Luengo-Navas, J., & Martos, J. M. (2017). Protestando en Twitter: Ciudadanía y empoderamiento desde la educación pública. [Protesting on Twitter: Citizenship and empowerment from public education]. *Comunicar: Revista Científica Iberoamericana de Comunicación y Educación*, 25(53), 39–48. <http://doi.org/10.3916/C53-2017-04>
- Scott, C. (2000). Accountability in the regulatory state. *Journal of Law and Society*, 27(1), 38–60. <https://doi.org/10.1111/1467-6478.00146>
- Serpieri, R., Grimaldi, E., & Vatrella, S. (2015). School evaluation and consultancy in Italy. Sliding doors towards privatisation? *Journal of Educational Administration and History*, 47(3), 294–314. <http://doi.org/10.1080/00220620.2015.1038695>
- Simmonds, M., & Webb, P. T. (2013). Accountability synopticism : How a think tank and the media developed a quasi- market for school choice in British Columbia. *International Education Journal: Comparative Perspectives*, 12(2), 21–41. Retrieved from <https://openjournals.library.sydney.edu.au/index.php/IEJ/article/view/7454>

- Simons, A., & Voß, J.P. (2018). The concept of instrument constituencies: Accounting for dynamics and practices of knowing governance. *Policy and Society*, 37(1), 14–35, doi: 10.1080/14494035.2017.1375248
- Skedsmo, G. (2011). Formulation and realisation of evaluation policy: Inconsistencies and problematic issues. *Educational Assessment, Evaluation and Accountability*, 23(1), 5–20. <http://doi.org/10.1007/s11092-010-9110-2>
- Solhaug, T. (2011). New public management in educational reform in Norway. *Policy Futures in Education*, 9(2), 267–279. <http://doi.org/10.2304/pfie.2011.9.2.267>
- Stamelos, G., Vassilopoulos, A., & Bartzakli, M. (2012). Understanding the difficulties of implementation of a teachers' evaluation system in Greek primary education: From national past to European influences. *European Educational Research Journal*, 11(4), 545–557. <http://doi.org/10.2304/eeerj.2012.11.4.545>
- Teltemann, J., & Jude, N. (2018, July). New accountability schemes? Assessing trends in educational assessment and accountability procedures in OECD countries. Paper presented at the XIX International Sociological Association (ISA) World Conference, Toronto, Canada.
- Termes, A., & Mentini, L. (2018, September). From low to high stakes: Ideational and material drivers behind test-based accountability reforms in Minas Gerais, Brazil. Paper presented at the European Conference on Educational Research (ECER), Bolzano, Italy.
- Thelen, K. (2004). How institutions evolve. In J. Mahoney, & D. Rueschemeyer (Eds.), *Comparative historical analysis in the social sciences* (pp. 208–240). Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9780511803963.007>
- Thiel, C., Schweizer, S., & Bellmann, J. (2017). Rethinking side effects of accountability in education: Insights from a multiple methods study in four German school systems. *Education Policy Analysis Archives*, 25(93). <http://doi.org/10.14507/epaa.25.2662>
- Tveit, S. (2009). Educational assessment in Norway – A time of change. In C. Wyatt-Smith, & J. J. Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory and practice* (pp. 227–243). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-1-4020-9964-9_12
- Vaillant, D. (2005). *Education reforms and teachers' unions: Avenues for action* (IIEP Fundamentals of educational planning 82). Paris, France: UNESCO–International Institute for Educational Planning. Retrieved from unesdoc.unesco.org/images/0014/001410/141028e.pdf

- Veloso, L., Abrantes, P., & Craveiro, D. (2013). The Portuguese schools' evaluation programme: A sociological approach to the participation of social actors. *Evaluation*, 19(2), 110–125. <http://doi.org/10.1177/1356389013485616>
- Verger, A., & Curran, M. (2014). New public management as a global education policy: Its adoption and re-contextualization in a Southern European setting. *Critical Studies in Education*, 55(3), 253–271. <http://doi.org/10.1080/17508487.2014.913531>
- Verger, A., Parcerisa, L., & Fontdevila, C. (2018). The growth and spread of large-scale assessments and test-based accountabilities: a political sociology of global education reforms. *Educational Review*, 71(1), 1-26. <https://doi.org/10.1080/00131911.2019.1522045>
- Vetterlein, A., & Moschella, M. (2014). International organizations and organizational fields: Explaining policy change in the IMF. *European Political Science Review*, 6(01), 143–165. <http://doi.org/10.1017/S175577391200029X>
- Wang, Y. (2017). The social networks and paradoxes of the opt-out movement amid the Common Core State Standards implementation: The case of New York. *Education Policy Analysis Archives*, 25(1), 1–27.
- Whetton, C. (2009). A brief history of a testing time: National curriculum assessment in England 1989–2008. *Educational Research*, 51(2), 137–159. <http://doi.org/10.1080/00131880902891222>
- Whitty, G., & Wisby, E. (2016). Education in England – A testbed for network governance? *Oxford Review of Education*, 42(3), 316–329. <http://doi.org/10.1080/03054985.2016.1184873>
- Wilkins, A., & Olmedo, A. (Eds.). (2019). *Education governance and social theory: Interdisciplinary approaches to research*. London: Bloomsbury Publishing.
- Ydesen, C. (2013). Educational testing as an accountability measure: Drawing on twentieth-century Danish history of education experiences. *Paedagogica Historica*, 49(5), 716–733. <http://doi.org/10.1080/00309230.2013.815235>
- Ydesen, C., & Andreasen, K. E. (2014). Accountability practices in the history of Danish primary public education from the 1660s to the present. *Education Policy Analysis Archives*, 22(120). <http://doi.org/10.14507/epaa.v22.1618>

Appendix: List of terms included in the search protocol

TITLE-ABSTRACT-KEYWORDS (("accountability" OR ((("high-stakes" OR "low-stakes") W/2 "test*")) OR "managerial*" OR "incentive" OR "sanction" OR "reward" OR "bonus" OR "performance-based pay" OR "pay-per-performance" OR "merit-based pay" OR "teacher evaluation" OR "feedback" OR "school evaluation" OR "school self-evaluation" OR "school inspection" OR "standardized test*") AND ("test*" OR "learning outcomes" OR "standard" OR "evaluat*" OR "rank*" OR "benchmark*" OR "assess*" OR "result" OR "grad*" OR ("performance" W/2 ("teacher" OR "student" OR "school"))))

AND TITLE-ABSTRACT-KEYWORDS ("International organization" OR "IO" OR "NGOs" OR "unions" OR "teachers organization" OR "aid agency" OR "international community" OR "civil society" OR "policy actor" OR "policy agent" OR "non-state actor" OR "global actor" OR "government" OR "international agency" OR "multi-lateral agency" OR "think tank" OR "stakeholder" OR "policy entrepreneur" OR "policy network" OR "policy maker" OR "decision maker" OR "practitioner" OR "state" OR "private sector" OR "corporate actor" OR "provider" OR "education business" OR "advocate" OR "coalition" OR "stakeholder" OR "education industry" OR "PISA" OR "OECD" OR "ILSA" OR "international large-scale assessment" OR "Political economy" OR "government" OR "governance" OR "policy trajectory*" OR "policy-maker*" OR "policy-shape*" OR "practitioner" OR "politics of education" OR "education policy*" OR "economics of education" OR "policy implementation" OR "policy borrowing" OR "policy lending" OR "agenda" OR "advocacy" OR "lobbying" OR "globalization" OR "globalizing world" OR "westernization" OR "Europeanization" OR "world society" OR "global arena" OR "multi-level govern*" OR "multi-scalar govern*" OR "decision maker*" OR "policy-practice" OR "institutionalism" OR "critical theory" OR "neoclassic economy*" OR "neoliberalism" OR "neo-liberalism" OR "human capital" OR "knowledge economy" OR "knowledge market" OR "policy entrepreneur" OR "policy networks" OR "policy paradigm" OR "policy learning" OR "policy convergence" OR "policy transfer" OR "policy travelling" OR "global education marketplace" OR "capitalism*" OR "global network" OR "policy network" OR "network governance" OR

"think tank" OR "policy communit*" OR "structures of power" OR "policy change" OR "education* reform")

AND TITLE-ABSTRACT-KEYWORDS ("school" OR "schooling" OR "education" OR "educational" OR "vocational training" OR "VET" OR "TVET" OR "professional training")

Publication B: Parcerisa, L., & Falabella, A. (2017). La consolidación del Estado evaluador a través de políticas de rendición de cuentas: trayectoria, producción y tensiones en el sistema educativo chileno [*The consolidation of the evaluative state through accountability policies: Trajectory, enactment and tensions in the Chilean education system*]. *Education Policy Analysis Archives*, (25), 1-24. <http://dx.doi.org/10.14507/epaa.25.3177>

CHAPTER 2

La Consolidación del Estado Evaluador a Través de Políticas de Rendición de Cuentas: Trayectoria, Producción y Tensiones en el Sistema Educativo Chileno

Introducción

En las últimas décadas, la política de Rendición de Cuentas (RdC) ha penetrado en una amplia variedad de países de todo el mundo. En el caso chileno, la expansión y consolidación de este enfoque surge a propósito de las críticas y “fallas” del mercado escolar, que se viene implementando desde los años 80s. Las reformas educativas pro-mercado que se adoptaron durante esa década, convirtieron a Chile en el primer país del mundo en implementar un sistema de subsidio a la demanda vía “voucher” de manera universal (Zancajo, 2017).

Paradójicamente, en un esquema altamente desregulado, un Estado fuerte, que estandariza, evalúa y sanciona, emerge como “el salvador” del modelo mercantil. La *mano invisible* del mercado no era suficiente para regular y promover una educación de calidad y, ante tal constatación, se opta por recurrir a la *mano visible del Estado* a través de la RdC (Falabella, 2015, en prensa). Ello convierte Chile en un caso particular ya que, como señalan Verger y Normand (2015), la tendencia internacional ha sido la de utilizar la RdC como herramienta para introducir dinámicas de mercado a los sistemas educativos, mientras que en Chile, la RdC emergió a posteriori del diseño mercantil, justamente para intentar corregir las fallas de un esquema altamente liberalizado.

En Chile, las políticas de mercado en educación se empezaron a aplicar hace casi cuarenta años y fueron impulsadas por la dictadura neoliberal (1973-1990), que llevó a cabo un amplio proceso de privatización. Las piedras angulares de este sistema son un sistema de voucher, la libertad de elección del centro escolar y un sistema nacional de evaluación de aprendizajes (SIMCE) cuyos resultados por establecimiento han sido difundidos públicamente entre los años 2000 (Fernández, 2007) y 2015 (Mineduc, 2016). Los defensores de este modelo confiaban en que la competencia entre proveedores mejoraría la calidad educativa. Sin embargo, su aplicación práctica ha derivado en la generación de múltiples alteraciones, como el casi nulo control sobre

el uso de los recursos estatales, la significativa segregación socio-educativa y la inequidad en la calidad de la oferta escolar (Bellei, 2015; Verger, Bonal, & Zancajo, 2016).

Desde los inicios del modelo, bajo la dictadura, se diseñó, de modo incipiente, un ‘Estado evaluador’, que direccionaba “a la distancia” una dispersa red de ofertantes públicos y privados. En este período, se legisló respecto a la evaluación y publicación del desempeño escolar para orientar la “libre” elección y salida de los padres y empujar a que los profesionales de la educación tuvieran que rendirles cuentas a la comunidad por su desempeño. También, como evidencia Falabella (en prensa), se aspiraba a un sistema con sanciones estatales cuando la “comunidad” no fuera capaz de presionar ante ofertas escolares de baja calidad. No obstante, no se alcanzaron a concretizar dichas medidas de difusión e intervención, a pesar de ser ideales que circulaban en la época.

Con la recuperación de la democracia, asume el poder ejecutivo la Concertación de Partidos Por la Democracia, conglomeración de centro-izquierda, por cuatro períodos presidenciales (1990-2010). Durante la década de los 90 los gobiernos de centro-izquierda mantuvieron el esquema de financiamiento vía voucher e incluso se sumó la posibilidad de que los establecimientos escolares pudiesen cobrar una cuota a los padres (“financiamiento compartido”). Este esquema se complementó con, lo que fue llamado por el Ministerio de Educación, un “Estado activo” para “compensar el mercado”, considerando que las escuelas habían quedado en condiciones paupérrimas posteriormente a la dictadura cívico-militar. Este enfoque implicó medidas, tales como: aumento del gasto fiscal, entrega de apoyo material y pedagógico a los establecimientos, capacitaciones a docentes y programas de discriminación positiva.

En este período se continuaron desarrollando instrumentos de evaluación estandarizada y se introdujeron las primeras medidas de RdC. En 1995 se comenzaron a difundir los resultados de las pruebas SIMCE en los medios (estipulado en la LOCE³, 1990) y se promulgó el Sistema Nacional de Evaluación del Desempeño (SNED), que establece un sistema de bonos salariales colectivos asociados al desempeño escolar.

³ Ley Orgánica Constitucional de la Enseñanza.

No obstante los avances anteriores, en la década de los 2000 es cuando se iniciaron con mayor fuerza las medidas de RdC y se ampliaron las evaluaciones estandarizadas. Este enfoque se articularía de manera más completa con la promulgación de la Ley de Subvención Escolar Preferencial (SEP, 2008), que otorga a las escuelas un mayor subsidio (voucher) por cada alumno proveniente de contextos de pobreza. Bajo este nuevo esquema, mientras los establecimientos pueden recibir mayor financiamiento, éstos deben rendir cuentas al Estado por el cumplimiento de metas e indicadores estandarizados. Además los establecimientos son clasificados según su desempeño y un bajo rendimiento conlleva a sanciones que incluye la posibilidad de cierre.

Posteriormente, se consolida el enfoque de RdC bajo el “Sistema de Aseguramiento de la Calidad de la Educación” (2011). Éste conlleva la instauración de una nueva matriz institucional que se encarga de fiscalizar, evaluar, inspeccionar y sancionar todos los establecimientos públicos y privados-subvencionados. De este modo, se consolida un “Estado evaluador”, con la expectativa de corregir las principales fallas y omisiones del modelo de mercado escolar instaurado desde los años 80.

El objetivo de este artículo es doble. En primer lugar, examinamos las razones que explican la adopción de las políticas de RdC en Chile y, particularmente, del Sistema de Aseguramiento de la Calidad de la Educación (en adelante SACE). En segundo lugar, analizamos el proceso de elaboración de la política SACE, prestando especial atención a las relaciones dinámicas entre contextos, actores y discursos a la hora de producir dicha política. El artículo ofrece una descripción densa de la trayectoria y del proceso de adopción y re-contextualización de las medidas de RdC implementadas durante los 2000, siguiendo el enfoque elaborado por Verger (2016) sobre la política y la semiótica de la adopción de políticas.

El artículo se estructura en cinco secciones. En la primera sección se presenta la RdC como política educativa global y se introducen los principales elementos teórico-analíticos que orientaran la investigación. Seguidamente, se presenta la estrategia metodológica que se ha seguido para elaborar el estudio. En tercer lugar, se expone cómo se construyeron las crisis del sistema educacional y cuáles fueron los factores que desencadenaron la apertura de una “ventana de oportunidad” (Kingdon, 1995) que facilitó la consolidación de la RdC en el esquema de mercado escolar. En cuarto lugar, se presenta el desarrollo del proceso de adopción de la política

y la emergencia de la RdC como lugar de consenso y ambigüedad entre el centro-izquierda y la derecha. Por último, se examinan las distintas racionalidades políticas en juego y las disputas discursivas durante el proceso de deliberación y promulgación de la política.

La Rendición de Cuentas como Política Educativa Global: Aportes Teórico-Analíticos para Investigar su Adopción y Re-Contextualización

La RdC es una pieza central en las recomendaciones elaboradas por los principales organismos internacionales en el campo educativo, como por ejemplo, la Organización para la Cooperación y el Desarrollo Económico (OCDE) y el Banco Mundial (BM). A este respecto, la RdC se considera central para mejorar los resultados (OCDE, 2009) y la calidad de los sistemas educativos, y para alinear los incentivos que orientan la acción de los distintos actores escolares (World Bank, 2015).

En el campo de la educación, la RdC incluye mecanismos y procesos destinados a asegurar el cumplimiento de los objetivos y las obligaciones de los actores, y especialmente (aunque no exclusivamente), de las escuelas, los directivos y los docentes (Hatch, 2013). Según Bovens (2007), la RdC se puede definir como una relación que se establece entre un actor (por ejemplo, la escuela o los docentes) y una contraparte (como el Estado, una institución o las familias). En esta relación, el actor en cuestión tiene el deber y la obligación de rendir cuentas sobre su comportamiento ante el foro, mientras que este último tiene la capacidad de formular preguntas, solicitar evidencias y emitir juicios, que pueden conllevar consecuencias para el actor.

Sin embargo, en el discurso político en educación, la RdC a menudo se emplea como un “concepto paraguas” (Gunter et al., 2016). La RdC puede adoptar múltiples configuraciones e incluye diferentes opciones de diseño y modelos, como por ejemplo, el modelo de mercado, el burocrático o el legal. Sin embargo, el modelo de RdC basada en pruebas (o gerencial) se está convirtiendo en el modelo hegemónico a nivel internacional (Parcerisa & Verger, 2016).

Las políticas de RdC son una herramienta clave para la creación de un “mercado performativo” (Falabella, 2014) y para la “endoprivatización” de los sistemas educativos (Ball & Youdell, 2008). Los mercados escolares performativos combinan la descentralización de la gestión escolar, la

liberalización del sistema educativo y la introducción de sistemas de financiamiento estatal basados en la demanda (o de “voucher”), con un Estado evaluador o “supervigilante” que se encarga de definir el currículum y estándares nacionales, así como también de evaluar las escuelas, difundir sus resultados y asignar las consecuencias que se asocian al desempeño logrado (que pueden tomar forma de recompensa, asesoría o de sanción).

En este contexto, el Estado aumenta su poder, convirtiéndose en un “Estado evaluador” (Maroy, 2009) que gobierna una dispersa red de instituciones educativas “a la distancia” (Rose & Miller, 1992). El Estado ejerce de intermediario y guía en la relación mercantil que se establece entre oferta y demanda. Por otra parte, la escuela está sujeta a una “doble rendición de cuentas”, frente al Estado (a quién deben mostrar el logro de los estándares de desempeño) y frente a la demanda (las familias, a quien deben mostrar una oferta atractiva para aumentar la matrícula del centro, y con ella, los recursos). Esta configuración sigue una lógica basada en la *Nueva Gestión Pública*, la cual abarca una amplia variedad de políticas, programas y dispositivos de reingeniería organizativa y cultural (ver Newman, 2005; Verger & Normand, 2015). Ello implica, como argumentan Ball y Youdell (2008), la transformación tanto de las estructuras, como de los valores y comportamientos de los trabajadores de la educación y del *ethos* de las instituciones escolares, aunque dicha transformación no está libre de tensiones.

Con el fin de analizar la adopción y re-contextualización de las políticas de RdC en Chile, y más concretamente su proceso de adopción y re-contextualización, utilizamos herramientas teórico-analíticas basadas en el enfoque de “la política y la semiótica de la adopción de políticas” (Verger, 2016), que combina elementos procedentes de dos perspectivas distintas: el Constructivismo Crítico (ver Hass, 2004; Hay, 2002) y la Economía Política Cultural (ver Jessop, 2010). Este enfoque es particularmente útil ya que permite entender el proceso de adopción como un proceso dialéctico, en el que intervienen ideas, discursos, relaciones de poder, así como conflictos y espacios de negociación entre actores, que se desarrollan en momentos y contextos concretos (Verger, 2016).

El Constructivismo Crítico, por una parte, considera que para entender el cambio político, debemos fijarnos en la relación dinámica que se da entre las ideas, los actores y los contextos selectivos que favorecen la adopción y el impacto de ciertas estrategias políticas, en detrimento

de otras (Hay, 2002). Este enfoque atribuye a las ideas el rol de variable independiente a la hora de analizar el cambio de la política (Verger, 2014). Por otro lado, la perspectiva de la Economía Política Cultural (EPC) aporta herramientas heurísticas que tienen una gran utilidad para la investigación sobre la producción de la política (Verger, 2014, 2016). Esta perspectiva subraya la importancia de los factores semióticos en el análisis de la articulación de las relaciones que se establecen entre la esfera política, económica y social (Jessop, 2010). Según Verger (2016), ambas perspectivas se pueden enriquecer mutuamente ya que comparten principios ontológicos.

A este respecto, Jessop (2010) plantea que existen tres mecanismos evolucionarios que permiten estudiar los procesos de adopción de las ideas globales: la variación, la selección y la retención. La variación es el punto de partida, un momento donde los discursos y políticas sedimentadas entran en crisis, abriéndose una grieta para su re-politización y cambio. Los momentos de variación de la política pueden favorecer la apertura de “ventanas de oportunidad” para que actores concretos promuevan sus demandas, llamando la atención sobre determinadas problemáticas o situando sus reivindicaciones en el debate público (Kingdon, 1995). En esta fase se generan múltiples interpretaciones sobre las causas y las posibles soluciones a la crisis que entran en competición. A la hora de analizar el momento de variación, cabe fijarse en factores de carácter histórico, político y económico, así como en las presiones ejercidas por múltiples actores (organismos internacionales, movimientos sociales, etc.) para legitimar la necesidad del cambio en la política.

La selección, implica la identificación, por parte de los *policy-makers*, de una narrativa particular sobre las causas y las soluciones más idóneas para resolver la crisis. El deseo de la política, que a menudo se fundamenta en la existencia o la percepción de un problema o una crisis que se debe resolver, se acostumbra a asociar a una solución, detrás de la cual se esconden asunciones sobre la teoría del cambio, ambigüedades, múltiples intereses en conflicto, omisiones y contradicciones. En el proceso de selección de la política, juegan un rol central aspectos como su credibilidad empírica y la disponibilidad de evidencias, las percepciones sobre su “viabilidad administrativa” (ver Hall, 1993) y la “ideología” de los tomadores de decisiones (Verger, 2014; Verger, Fontdevila & Zancajo, 2016).

La fase de retención se refiere al proceso de sedimentación e institucionalización de los discursos en programas políticos concretos, y toman especial relevancia las contingencias locales (ver Ball, 1998; Campbell, 2004), principalmente de carácter material. Por otro lado, la retención no es ajena a los conflictos latentes y manifiestos existentes durante el proceso de formulación y producción de la política, las estrategias desarrolladas por los actores y la correlación de fuerzas existente a nivel parlamentario, etc.

En definitiva, se trata de analizar las interacciones entre la dimensión contextual y material en la que se formula y se produce la política, con la dimensión semiótica, en la que se producen y circulan las visiones y comprensiones de los actores (Verger, Curran & Parcerisa, 2015). En este sentido, los tres momentos identificados por Jessop (2010) y desarrollados por Verger, Fontdevila y Zancajo (2016), y Verger (2014, 2016), no pueden entenderse como momentos separados unos de otros, sino como momentos que se relacionan dialécticamente y que sirven como estrategia empírica-analítica para la investigación de la adopción de políticas.

Estrategia Metodológica

El estudio pretende analizar los discursos oficiales y el proceso de producción de las políticas de accountability en educación en Chile, así como examinar los actores involucrados y los factores (políticos, institucionales, culturales y económicos) que han mediado en la adopción de estas reformas. A nivel metodológico, el estudio combina entrevistas semi-estructuradas con análisis de documentos. En total se realizaron 27 entrevistas, a actores involucrados en el debate y construcción de la política de “aseguramiento de la calidad” entre los años 2006 - 2011. La selección de los entrevistados fue intencional y se basó en dos ejes: la ideología y el rol del actor en el proceso de adopción de la reforma, distinguiendo entre políticos, funcionarios públicos, expertos y representantes de actores sociales (centros de pensamiento, organismos internacionales, movimientos sociales, sindicatos, etc.). Como resultado se entrevistó: directivos y funcionarios del Mineduc (5), miembros de la Agencia de Calidad de la Educación (1), asesores del Mineduc (4), parlamentarios de derecha y centro-izquierda (2), representantes de centros de pensamiento de derecha (2), funcionarios de organismos internacionales (2), líderes del movimiento estudiantil (2), miembros del movimiento magisterial (3), centros de pensamiento de izquierda (2) y académicos que participaron en el proceso de reforma educacional (4).

En el análisis documental se incluyen leyes (Ley de Subvención Escolar Preferencial, Ley General de Educación, Ley de Sistema Nacional de Aseguramiento de la Calidad de la Educación) y actas parlamentarias, informes gubernamentales (Informe final del Consejo Asesor Presidencial para la Calidad de la Educación), informes de organismos internacionales (OCDE, Banco Mundial), cuentas públicas anuales de los Presidentes, discursos de los Ministros de Educación de inauguración del año escolar y documentos públicos de los actores sociales, entre otros. En total se revisaron aproximadamente 4000 páginas de documentos oficiales.

Para examinar el material cualitativo se procedió a realizar análisis crítico del discurso (Fairclough, 1995), y se codificó el total del material a partir de códigos pre-establecidos referentes a la identificación de componentes claves del proceso de adopción de la política (rol del mercado, el rol del Estado, autonomía escolar, rendición de cuentas, educación pública y privada, entre otros). Las entrevistas fueron transcritas manualmente y se procedió al análisis mediante el software de análisis cualitativo Atlas.ti, versión 6.

La Producción de la Crisis y la Ventana de Oportunidad para la Reforma Educativa

Las crisis en la política se construyen discursivamente como una urgencia mediante la combinación de sucesos, datos, emociones, y son elevadas a la categoría de situaciones altamente problemáticas. Entonces, se crea la justificación de cambio de las estructuras, narrativas y prácticas vigentes en el sistema educacional (Verger, 2016).

En el caso del fortalecimiento de las políticas de RdC en Chile hubo dos momentos de “variación” (Jessop, 2010) que se comprenden como momentos críticos y se utilizaron para justificar la intensificación de la política de RdC⁴. La primera crisis ocurre entre fines de la década de los 90s y principios de los 2000 referida al “estancamiento” e inequidad en los resultados nacionales (SIMCE). A lo cual, luego se suma el deficiente desempeño en pruebas

⁴ Desde la reforma de los años 80 otros momentos de percepción de crisis fueron claves para la adopción de las políticas. A finales de los 70 se creó la idea de una crisis de la educación pública, desfinanciada, ineficiente y de baja calidad; luego a principios de los 90, la crisis era referida al abandono del Estado durante la dictadura.

internacionales, como por ejemplo, el *Programa Internacional de Evaluación de Estudiantes* (PISA, por sus siglas en inglés) y el *Estudio Internacional de Tendencias en Matemáticas y Ciencias* (TIMSS) (ver Bellei & Vanni, 2015). Posteriormente en el año 2006, emerge el segundo momento de quiebre, la “revolución pingüina”, que significó masivas protestas estudiantiles en contra del modelo escolar impuesto desde los años 80. Ambas crisis abrieron una “ventana de oportunidad” (Kingdon, 1995), que facilitó la introducción de las políticas de rendición de cuentas en el sistema educativo.

La primera “crisis” sucedió después de una década de inversiones y reformas en el sector educativo, como relata un funcionario del Ministerio de Educación de la época:

Y aparecen los resultados del SIMCE de 4° básico el año 2001 [aplicado el año 2000] ... Y esta era una generación de 4° básico que como que se consideraba que eran los que habían partido en la escuela con la reforma ya en marcha ... es como que había pasado o había estado, había sido parte de la reforma... y los resultados que aparecieron fueron no muy optimistas, o muy lejos de las expectativas, en términos de que fueron unos resultados que comparados con la generación de 4° básico que lo daba el año '96, eran más o menos similares. Eso empieza a generar un super fuerte debate respecto de... de la reforma y de su legitimidad y si lo estamos haciendo bien. Y de que los actores políticos más de... los partidos políticos, los técnicos más de oposición, de derechas en ese momento, empiezan a dispararle a las iniciativas de políticas de reforma con mucha mayor fuerza. El mundo político en general, incluso desde los partidos del Gobierno, empiezan a mirar, a desconfiar de lo que se estaba haciendo porque no se lograban resultados. Tiempo después aparecen... Chile empieza a participar en algunas pruebas internacionales, aparecen los resultados del TIMSS donde Chile aparece con tampoco no muy buenos resultados en el contexto internacional. Entonces, todo eso como que pone una presión super fuerte respecto de... pucha, se están haciendo muchas cosas pero parece que no están haciendo efecto, o como el efecto esperado por lo menos. (Entrevista Exdirectivo Mineduc01)

La cita anterior evidencia la presión política⁵ y mediática (particularmente desde el diario “El Mercurio”) que estaba recibiendo el gobierno en ese entonces, y cómo los bajos resultados estandarizados ponían en cuestión la eficacia de la política del “Estado activo”.

La decisión de la Concertación de publicar los resultados SIMCE anualmente y participar de las pruebas internacionales, generó un ambiente político que le dio relevancia y legitimidad a dichos

⁵ Según Bellei y Vanni (2015), la oposición de derecha argumentaba que el “estancamiento de los resultados” era debido a la escasa eficacia mostrada por las políticas educacionales de corte estatista que se implementaron durante ese período (por ejemplo, una crítica central apunta a las “rigideces laborales” inherentes al Estatuto Docente).

resultados como una medición válida a las políticas educativas. Por ende, el primero en estar obligado en demostrar su eficacia por medio de mediciones, es el propio gobierno.

Como consecuencia, la relevancia y la veracidad otorgada a los resultados del SIMCE provocó impulsar políticas alineadas con el sistema de medición que buscasen incrementar los resultados, y por ende el deseo de traspasar la presión hacia los establecimientos.

Desde el punto de vista del Ministerio, era como, pucha, tenemos un problema, tenemos programas de apoyo a las escuelas, programas de apoyo focalizados en las escuelas más pobres o las que tienen más débiles resultados, pero tienen unos efectos bien limitados. No logramos con eso cambiar mucho, por ejemplo, P900 y otras cosas y otros programas no logran... y la sensación de que es un Ministerio que como no administra las escuelas directamente, los administran los municipios o los privados, no tiene como... la frase coloquial que se acuñó es que era un Ministerio “sin dientes”. En el sentido que podía generar nuevas orientaciones, podía generar nuevos recursos, podía generar materiales, programas de apoyo, pero no podía hacer mucho porque no tenía muchos instrumentos para poder ejercer mucha presión [...] (Entrevista Exdirectivo Mineduc01).

Las entrevistas reflejan que al interior del Ministerio había una *sensación de despoder*, un Ministerio “sin dientes”, por carecer de mecanismos concretos para presionar a los establecimientos. Esta sensación aumenta en un contexto con un mercado escolar altamente desregulado, donde los requisitos para la apertura de nuevos establecimientos educacionales eran escasos, y los proveedores privados representaban casi un 50% de la matrícula total. Las escuelas con bajos resultados, expresan los entrevistados, disfrutaban de “total impunidad” y seguían recibiendo una subvención procedente de fondos públicos sin ninguna exigencia relacionada con la “calidad educativa” ofrecida a los estudiantes y a sus familias. Por ende, estos discursos perciben como problemático que el Estado no tenga ni la capacidad ni las atribuciones necesarias para ejercer presión en las escuelas e influir en sus prácticas.

De esta manera, la sensación del despoder, junto a la aspiración de mayor control estatal sobre el mercado educativo, se potencia con la presión mediática hacia el gobierno de demostrar resultados evaluativos. La lógica detrás de esta idea es que el apoyo y los recursos del Estado son condición necesaria pero no suficiente para alinear los intereses, generar cambios en los comportamientos de los actores (establecimientos educacionales, sostenedores, directivos, docentes, etc.) y en las prácticas a nivel de aula, que deben conducir al mejoramiento de la calidad.

Paralelamente, a principios de la década de los 2000, organismos internacionales como la Organización para la Cooperación y el Desarrollo Económico también señalan el excesivo peso del mercado en el sistema educativo chileno y la necesidad de regulación del mismo. En este sentido, en la “*Revisión de políticas nacionales de educación*” de Chile, la OCDE afirma que “la educación chilena está influenciada por una ideología que da una importancia indebida a los mecanismos de mercado para mejorar la enseñanza y el aprendizaje” (OCDE, 2004, p. 290).

Progresivamente, se sedimentó la idea de que las reformas educativas de los 90 no habían logrado “llegar al aula” y la necesidad de consolidar un “Estado evaluador” (Santa Cruz, 2016). Se comienza a dibujar de manera más clara la figura del Estado evaluador bajo una lógica gerencial que orienta las escuelas a “demostrar” el logro de las metas e indicadores fijados por el Estado y un desempeño de “calidad”. Demostrar, evidenciar, hacer observable el buen desempeño empieza a tornarse un eje de la gestión escolar. Por ende, se *fracciona la responsabilidad*, si antes era el gobierno el responsable de dar cuenta del “estancamiento de resultados”, ahora debían ser las escuelas.

En consecuencia, se empezaron a establecer sistemas de incentivos individuales, como la Asignación de Excelencia Pedagógica (2002), que es un programa de incentivos económicos de carácter voluntario, o la Evaluación Docente (2004), que se aplica a los docentes ocupados en el sector municipal (Mizala & Schneider, 2014).

Es importante notar que, según el discurso oficial, la introducción de las medidas de RdC (efecto) son consecuencia de una “crisis de resultados” (la causa), en este análisis se argumenta que sucede más bien lo contrario. Así, *es la política la que creó la crisis*, pues primero se valora y decide medir y publicar los resultados como algo deseable y creíble, y luego la propia política de evaluación (el SIMCE) crea la necesidad de más política de RdC.

La segunda crisis se relaciona con el malestar social existente entre la sociedad civil derivado de las políticas educativas orientadas hacia el mercado que se habían implementado, y sus efectos sobre la calidad y la equidad. Las políticas neoliberales habían contribuido a la creación de un sistema escolar altamente segregado y con importantes niveles de desigualdad (Valenzuela, Bellei

& De Los Ríos, 2010, 2014), que facilitó el surgimiento de un malestar social latente que cristalizó en una amplia protesta social liderada por los estudiantes secundarios. En suma, lo que distingue esta crisis de la anterior es su origen “*bottom-up*” y la introducción de la idea de equidad/justicia en el debate público.

El movimiento de estudiantes secundarios de 2006 emergió en primera instancia como respuesta al aumento del costo de la prueba de acceso a la universidad y del transporte. Sin embargo, el descontento social producido por estos hechos rápidamente derivó en una respuesta frente a las crecientes desigualdades generadas por el mercado educativo y las diferencias en la calidad de la educación ofrecida por las distintas tipologías de establecimientos educacionales (Bellei & Cabalín, 2013; Donoso, 2013).

Durante su apogeo, este movimiento social logró movilizar a una amplia mayoría de los estudiantes. Además, los estudiantes secundarios socializaron y generalizaron la percepción de crisis de la educación entre la sociedad civil y lograron situar la educación en el centro de la agenda política. En el programa político del primer Gobierno de Bachelet la educación era un tema secundario, pero frente a las masivas movilizaciones sociales (que combinaron diferentes repertorios de acción colectiva, como por ejemplo, manifestaciones y “tomas” de colegios), que contaban con un alto apoyo y legitimidad entre la opinión pública, la educación se convirtió en un tema prioritario.

En este contexto emergió una narrativa alternativa que fue articulada principalmente por el movimiento de estudiantes secundarios y otros aliados organizados alrededor del denominado Bloque Social, que incorporaba a estudiantes secundarios y posteriormente universitarios, docentes y apoderados (Colegio de Profesores, Confederación de Estudiantes de Chile, Confederación de Estudiantes de Educación Superior Privada y la Asociación de Apoderados, entre otros). Los estudiantes secundarios articularon un discurso que combinaba demandas concretas a corto plazo, con objetivos más ambiciosos vinculados con la defensa de la educación pública, el fin del mercado en educación y las prácticas discriminatorias (Bellei & Cabalín, 2013). En consecuencia, exigían cambios profundos en lo que se refiere a la arquitectura del sistema educativo. Esta narrativa significaba una crítica a las políticas neoliberales implementadas desde la dictadura, y relacionaba los problemas de calidad y equidad del sistema educativo con la

mercantilización de la educación, el rol subsidiario del Estado y el debilitamiento de la educación pública, que ha generado un sistema profundamente desigual:

[la educación] se convirtió en una mercancía, sometida a los estrechos criterios de eficiencia y competitividad, donde el Estado asume sólo un rol subsidiario y débilmente regulador, cuyo objetivo es financiar la demanda por educación pero no responsabilizarse por las reales necesidades educativas de la población. Esto ha generado una profunda segmentación y fragmentación del sistema, afectando gravemente su equidad y calidad y aumentando el desequilibrio socioeducativo en desmedro de los sectores más postergados de nuestra sociedad (Bloque Social, 2006, p. 4).

Frente a la crisis desencadenada por el movimiento de estudiantes secundarios, el gobierno creó un Consejo Asesor Presidencial para la Calidad de la Educación, que permitiría institucionalizar el conflicto social y abrir un espacio de consulta colectiva. El Consejo Asesor Presidencial fue conformado por 81 integrantes de distintas posiciones e incorporó a los representantes del movimiento estudiantil (Burton, 2012; Cox, 2012). Como veremos a continuación, en el Consejo Asesor Presidencial se identificaron distintas narrativas sobre las causas de la crisis del sistema educacional, y las potenciales soluciones a esta, sin embargo, la RdC, y particularmente el Sistema de Aseguramiento de la Calidad de la Educación (SACE), emergió como solución de consenso entre la centro-izquierda y la derecha.

El Estado Evaluador y la Rendición de Cuentas como Solución del Fallido Mercado Escolar Chileno

La segunda etapa o “momento” del proceso de adopción de una reforma se refiere a la “selección”. Se trata de un momento clave para comprender el proceso de adopción y cambio de las políticas educativas vigentes y se caracteriza por la identificación y selección estratégica, por parte de los *policy-makers*, de aquellos discursos sobre las causas de los problemas que existen en una sociedad (o en un área de la política en particular) y las soluciones para resolverlos (Verger, 2016). En el marco del “Consejo Asesor Presidencial para la Calidad de la Educación”, los diferentes actores que lo integraban intentaron influir en el proceso político imponiendo su relato sobre la crisis del sistema educativo y sus soluciones, mientras que el Mineduc quedó relegado a un segundo plano, dejando que otros actores lideraran la configuración de la agenda política.

Asimismo, el debate que se desarrolló en el interior del Consejo fue muy complejo y mostró la dificultad de lograr acuerdos en algunos aspectos particulares de las distintas “coaliciones programáticas”: izquierda, centro-izquierda (Concertación) y la derecha (Cox, 2012; Santa Cruz, 2016). A pesar de significativas diferencias entre las distintas coaliciones programáticas, el Informe final del Consejo Asesor Presidencial para la Calidad de la Educación (2006) plantea un gran consenso respecto a la necesidad de “definir estándares” y crear una “agencia pública de aseguramiento de la calidad”.

Es imprescindible definir estándares de calidad que todos los establecimientos del país satisfagan, proponer requisitos de entrada más exigentes a la actividad educativa y crear una Agencia Pública de Aseguramiento de la Calidad (Consejo Asesor Presidencial para la Calidad de la Educación, 2006, p. 15).

Poco a poco, iba tomando fuerza la idea de adoptar nuevos mecanismos de rendición de cuentas. En primer lugar, un sector de la centro-izquierda consideraba que la RdC permitiría fortalecer el rol “vigilante” del Estado, otorgándole una mayor capacidad de regular el funcionamiento del mercado y garantizar el derecho a una educación de calidad. En segundo lugar, la derecha interpretaba que la RdC era una herramienta útil para salvaguardar el mercado escolar, ya que encaja con los principios gerenciales de la Nueva Gestión Pública, donde los mecanismos de mercado y la orientación al cliente (esto es, la libertad de elección de las familias) juegan un rol prominente. Éste era un lugar de consenso, además, entre las élites intelectuales de la coalición de derecha (ver Beyer, 2001; Fontaine & Eyzaguirre, 2001) y de centro-izquierda (ver Traverso, 2004; Montt et al., 2006).

Sin embargo, los docentes y estudiantes decidieron salir del Consejo Asesor y desmarcarse de dicho informe argumentando que no se sentían representados por los contenidos del mismo (Vera, 2011). De esta manera, la coalición de izquierda demandaba profundizar en el diálogo ciudadano y se mostraba crítica con aquellas visiones de la educación de corte economicista presentes en el Consejo (Bloque Social, 2006), así como también con el redactado final del informe que no muestra con claridad quién defiende qué, y cuáles son las posiciones hegemónicas y subalternas en el interior del consejo (Santa Cruz, 2016).

Después de entregar el informe del Consejo Asesor Presidencial, el Gobierno tomó la iniciativa en el desarrollo de una nueva ley educativa que debía sustituir a la ley de Pinochet, la LOCE.

Este proyecto se acompañaría de otro proyecto de ley para la creación de una Superintendencia de Educación. El proyecto de Ley General de Educación (LGE) elaborado por la Concertación recogía algunos de los marcos discursivos y propuestas que contenía el informe que no generaron controversias significativas, como por ejemplo, el deber del Estado de asegurar a todos los estudiantes una educación de calidad, una educación obligatoria de 14 años, la promoción de espacios de participación en el sistema escolar, etc. Sin embargo, otros aspectos (como por ejemplo, el fin del lucro o la prohibición de seleccionar el alumnado hasta 8º año de educación general básica) generaron una fuerte controversia con la oposición de derecha (Alianza por Chile) y también con los sostenedores privados, a los que se unieron la Iglesia Católica y los sectores más liberales dentro de la Concertación. Además, las narrativas críticas con el proyecto de ley fueron amplificadas por los principales medios de prensa escrita, como por ejemplo, El Mercurio y la Tercera. Con esta oposición, la viabilidad política de la reforma era más que dudosa. En efecto, algunos de los artículos de la ley necesitaban un quórum calificado para su aprobación, y en una coyuntura donde la coalición de centro-izquierda en el gobierno (la Concertación) ya no poseía la mayoría parlamentaria, junto con el anuncio por parte de la oposición de su voto contrario, las posibilidades de aprobación de la ley eran remotas.

El rechazo al proyecto de ley expresado por sectores conservadores, junto con la caída de la popularidad del Gobierno (debido a una reforma en el transporte público de la ciudad de Santiago) y el veto de la derecha y de un sector liberal del Gobierno, debilitó al Gobierno y, con ello, dificultó el impulso de aquellas medidas que suscitaban más controversias. Por su parte, la derecha presentó su propia propuesta, articulada en una LGE alternativa que fortalecía los principios de autonomía escolar y rendición de cuentas, responsabilizando docentes y escuelas por el desempeño (con incentivos y sanciones) y desarrollando un Sistema de Aseguramiento de la Calidad (con una Superintendencia). A su vez, mantenía el rol subsidiario del Estado, aumentaba la transparencia de la información y fortalecía el derecho de elección de las familias. Este bloqueo, debido a la situación de empate en la correlación de fuerzas a nivel parlamentario, inclinó al Gobierno de la Concertación a tomar la decisión de construir un acuerdo con la oposición (Santa Cruz, 2016).

En este escenario, se decidió crear una comisión técnico-política formada por un grupo muy reducido de representantes políticos y asesores del gobierno de la Concertación y de la oposición

(Alianza por Chile) (ver: Larroulet & Montt, 2010). De esta manera, lo que había empezado siendo un proceso de deliberación que integraba a un amplio abanico de actores políticos y sociales, terminó siendo un debate cerrado con la participación exclusiva de una élite política (del Gobierno y de los partidos representados en el parlamento), encargada de construir los nuevos consensos en el campo de la educación (Burton, 2012). Esto refleja la tendencia al “duopolio” en los procesos de elaboración de políticas en Chile:

La historia es la siguiente. El gobierno de la época presentó un proyecto de ley, de reforma, la Ley General de Educación. La oposición de la época consideró que este no era un buen proyecto de ley y le pidió a un grupo de expertos que elaborará una alternativa. [...] Entonces, hicimos una propuesta alternativa. Y se armó un acuerdo político para tratar de compaginar y compatibilizar ambas propuestas. La negociación fue una negociación técnica política. Siempre con el norte de mejorar la calidad. Por eso que el tema de accountability no fue un tema de grandes diferencias. (Entrevista Exdirectivo Centro de Pensamiento 06)

La comisión técnica derivó en el “Acuerdo por la Calidad de la Educación” que se firmó en el año 2007. El acuerdo estableció las bases para “el Sistema de Aseguramiento de la Calidad” y, por ende, la consolidación del Estado evaluador y la política de RdC. Este acuerdo fue firmado por la Ministra de Educación Yasna Provoste, junto al Ministro de Secretaría General de la Presidencia y los presidentes de las principales fuerzas políticas de la coalición de centro-izquierda del Gobierno de la Concertación (Partido Socialista, Partido por la Democracia, Partido Radical, Democracia Cristiana) y de los partidos de la oposición (Renovación Nacional, Unión Demócrata Independiente). Como se ha dicho, a diferencia del alcance participativo que tuvo el Consejo Asesor Presidencial, la participación en esta comisión técnico-política fue muy reducida. Las conversaciones se llevaron a cabo con la más absoluta discreción y los principales debates de fondo se tuvieron en este espacio de deliberación.

En este contexto los mecanismos de rendición de cuentas (o “accountability”), enmarcados bajo una lógica de NGP, emergen como solución de consenso entre los políticos y legisladores, ya que permiten alinear sus preferencias ideológicas. En otras palabras, la RdC permitía mantener la arquitectura del mercado escolar, y a la vez corregir sus “fallas”. La pérdida de confianza en “la mano invisible” del mercado tuvo como consecuencia ser reemplazada por una “mano visible” del Estado, que adopta un rol regulador, evaluador y sancionador. La nueva racionalidad política dominante concibe de este modo al Estado como un vigilante y “árbitro” del mercado

escolar, que debe proteger “el derecho de los consumidores” (Falabella, 2015). Como resultado, lejos de revertir la lógica de mercado, las reformas de rendición de cuentas, materializadas en la Ley General de Educación (2009), y particularmente en el “Sistema de Aseguramiento de la Calidad” (2011), contribuyeron a sofisticar su funcionamiento.

Hay propuestas, pero yo diría que la gracia de esto es que nos pusimos de acuerdo las dos coaliciones políticas para buscar un consenso. Porque si tú te fijas, hay más rol del Estado, hay mayor, se crea la Agencia de Calidad de la Educación, se fomenta la obligación del Estado de dar información, pero eso facilita también la toma de decisiones por parte de las familias. Entonces, por eso que, te fijas, está construido para mejorar la calidad con un sistema mixto que satisfaga los requerimientos políticos de todos los sectores. (Entrevista Exdirectivo Centro de Pensamiento 06).

La RdC es una política *seductora* de capacidad *elástica*, argumenta Falabella (en prensa). La ambigüedad política y su mixtura de principios éticos (derecho/libertad; regulaciones/diversidad; público/privado) permitió conseguir el beneplácito de los sectores de derecha y de centroizquierda. Es justamente la narrativa dual la que conquista las hebras identitarias del “duopolio” chileno.

Intensificar la evaluación estandarizada, crear un sistema de inspección escolar y crear consecuencias según el nivel de desempeño escolar era una fórmula que circulaba entre los organismos internacionales como una “buena práctica” de política educativa (ver por ejemplo, Banco Mundial, 2007). Además, desde finales de los 90, se habían llevado a cabo prácticas de “turismo político” (Whitty, 2012), que tomaban forma de misiones en países anglosajones donde se había observado con atención los sistemas de aseguramiento de la calidad en educación. De este modo, países como Nueva Zelanda y el Reino Unido se identificaron como “sociedades de referencia” (Schriewer, 1990) para inspirar la reforma.

Era necesario establecer todo un mecanismo de contención y de regulación, de autonomía disciplinada, que en la nomenclatura del Banco Mundial se llama este modelo a la neozelandesa de concesiones por la calidad. En el fondo, centros escolares que se comprometen con un contrato con el Estado que haga cumplir, y en la medida que cumplan eso el Estado les otorga financiamiento, o uno más duro como el Reino Unido, donde además de contratos por la calidad, existe la idea de un Estado que interviene cuando hay dificultades, cuando las cosas no caminan, que exige... y que también apoya. (Entrevista Exdirectivo Mineduc 20)

Los factores anteriores vinculados al ambiente ideacional y al “soft power” (Verger, 2016), junto con otros factores de naturaleza política y económica fueron clave para la selección y adopción del SACE. En relación a estos últimos, se eligió además una opción pragmática frente a la crisis. En primer lugar, la existencia de capacidad de veto por parte de la derecha dificultaba la posibilidad de llevar a cabo cambios más profundos, tal y como se mostró con el rechazo del primer proyecto de LGE y de Superintendencia. En consecuencia, algunos temas fueron descartados porque se percibía que no generarían consenso, o bien, porque tenían un elevado coste económico y/o podrían abrir un nuevo frente de conflicto (como por ejemplo, el tema de las capacidades docentes).

Cambiar las dinámicas de mercado es difícil. Había poca disposición a eso. Poco piso político, poco consenso político para cambiar... y llegar a un acuerdo. Poco apoyo en el Congreso para hacer cambios estructurales en cuestiones como esa. En cambio, las cuestiones sobre atribuciones del Estado, sobre más poder, influencia, más accountability, había un poco más de consenso y parecían entonces por ende más viables políticamente y más viables de implementar. Y en el tema de las capacidades docentes lo que yo te decía cuando se discutió. Es como... pucha, es difícil y no sabemos mucho como hacerlo, es super caro, los programas que... hubo programas de apoyo a los docentes, de formación inicial y varias cosas en los 2000... (Entrevista Exdirectivo Mineduc 01)

Por último, uno de los factores cruciales para su selección fue la viabilidad económica de estas políticas, como narran algunos entrevistados, funcionarios del gobierno en aquel período. A diferencia de otras medidas, como por ejemplo el fin del lucro y del copago, estas no requerían una gran inversión monetaria, y por ende, facilitaba que el Ministerio de Hacienda apoyara la reforma. En definitiva, las razones y los factores que explican la selección de la RdC, y del SACE fueron múltiples, y combinan la convicción política, el pragmatismo, y el ambiente ideacional a nivel global y nacional.

La Construcción del Sistema Nacional de Aseguramiento de la Calidad: Fase de Institucionalización y Disputas de la Política

La fase de retención se refiere al momento en el que las soluciones políticas seleccionadas por los gobiernos son (o no) institucionalizadas. En esta fase tienen un gran peso los factores domésticos y las dinámicas de negociación y resistencia entre los principales actores (Verger, 2016). Para entender el proceso de selección, promulgación e institucionalización del SACE es

importante tener en cuenta la configuración del sistema político chileno de corte Presidencialista, y la capacidad de veto que tenía la derecha en ese momento, derivada de la correlación de fuerzas existente a nivel parlamentario.

El proceso de “institucionalización” del SACE se desarrolló mediante la promulgación de la Ley General de Educación (2009), donde se aprueba la nueva institucionalidad, y posteriormente la ley del Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media (2011), donde se desarrolla normativamente el SACE y su despliegue. Sin embargo, las negociaciones pre-legislativas que se llevaron a cabo en el marco de la comisión técnico-política fueron cruciales para comprender la forma que adoptaría finalmente el SACE.

En general, había un consenso entre el Gobierno y la oposición en el diagnóstico y en la solución, esto es, la necesidad de regular el mercado escolar para corregir sus fallas y alinear los incentivos de los actores, creando mecanismos de fiscalización y de control para presionar a las escuelas hacia la mejora de los resultados escolares. Sin embargo, a pesar del consenso que se había fraguado en la comisión técnico-política, durante el proceso de promulgación de la LGE hubo intensos debates y fuertes divergencias, especialmente dentro de la coalición de centro-izquierda. Dentro de la Concertación existían diferentes sectores con posiciones que no coincidían en cuestiones clave, como por ejemplo, el lucro, la selección de estudiantes y el rol de la educación pública.

En los debates parlamentarios que se desarrollaron en el marco de la discusión de la LGE (2009) y de la ley del SACE (2011), se vislumbró la estrechez de los desacuerdos entre ambas coaliciones respecto de la nueva institucionalidad y el fortalecimiento de la RdC (aunque respecto a otras políticas propuestas había mayores diferencias). En base al análisis de la discusión parlamentaria, se evidencia que durante la promulgación de la LGE, la mayoría de los parlamentarios se mostraron favorables a reforzar las políticas de RdC (también ver: Santa Cruz, 2016). Por otra parte, la evidencia que proporcionó el Banco Mundial por encargo del Ministerio de Educación, fue muy relevante en el debate de ambas leyes (Entrevista Parlamentario Concertación¹⁷) y resultó clave para “cambiar el veto que tenía la derecha respecto a la LOCE” (Entrevista Exdirectivo Centro de Pensamiento 05). El Banco Mundial se encargó de producir varios informes de diagnóstico y diseño del SACE que estuvieron muy presentes en el debate

parlamentario, sobre todo el primero titulado *El diseño institucional de un sistema efectivo de aseguramiento de la calidad de la educación en Chile* (Banco Mundial, 2007).

En el año 2008, las divisiones internas dentro del Gobierno de la Concertación hicieron que éste valorara la posibilidad de pactar con la derecha, ya que no disponía de los apoyos suficientes a nivel parlamentario para promulgar la reforma. Los sectores más izquierdistas de la coalición de Gobierno percibieron que el “protocolo de acuerdo” sobre la LGE se parecía más al contraproyecto de la derecha, que al proyecto de ley inicial del Gobierno (Santa Cruz, 2016).

En este contexto, el Gobierno decidió otorgar carácter de urgencia a la LGE, a pesar de la oposición del Colegio de Profesores, los estudiantes y una fracción importante de parlamentarios críticos de la propia coalición de centro-izquierda, que demandaban llevar a cabo un proceso deliberativo más profundo. De esta manera, el Gobierno logró censurar el debate y la aprobación de la ley, gracias a los votos favorables de la derecha. No obstante, la aprobación de la LGE se hizo en una coyuntura de fuertes divisiones internas dentro del Gobierno y de movilizaciones sociales lideradas por los docentes.

De hecho, durante la discusión en la sala, pocas semanas antes de la aprobación de la LGE, varios parlamentarios representantes del fragmento más crítico dentro de la Concertación manifestaban sus desacuerdos con la norma que se iba a promulgar, ya que consideraban que ésta no alteraba la lógica del mercado y no se comprometía firmemente con la educación pública y la lucha contra las desigualdades. Por ejemplo, la siguiente cita del Parlamentario Carlos Montes (Partido Socialista), ilustra el posicionamiento de esta fracción durante la discusión de la LGE:

la LGE contiene avances parciales [...], pero, en el fondo, no sale de la lógica del modelo de mercado y se mantienen sin modificación las cuestiones fundamentales. No hay una opción contundente por el humanismo en la educación; no se reconoce la importancia de la educación pública para enfrentar las desigualdades y para fomentar la integración, y no se establece la responsabilidad preferente del Estado por la educación pública (Carlos Montes, Historia de la Ley N° 20.370, p. 1498).

Posteriormente, el Gobierno de la Concertación enviaría el proyecto de ley del SACE, que fue aprobado en el Senado, aunque luego no superó el trámite en la Cámara de Diputados, que lo

rechazó y lo envió a Comisión Mixta⁶ (García-Huidobro & Bellei, 2010). En un contexto donde se acercaba el final de la legislatura del Gobierno de Bachelet, y siendo conscientes de que posiblemente la derecha ganaría las próximas elecciones y se encargaría de promulgar la Ley SACE, el sector crítico de la Concertación prefirió enviar el proyecto de ley a Comisión Mixta. El sector crítico creía que de este modo sería más factible influir en la definición final del SACE y corregir algunos de los “defectos” vinculados a los aspectos controvertidos enunciados anteriormente, a los que se sumaron la definición del concepto de calidad, el rol que debía desempeñar el Ministerio de Educación, las competencias atribuidas a la Superintendencia o el alineamiento del SACE con otras políticas de RdC como la ley de Subvención Escolar Preferencial (Berner & Bellei, 2011).

Yo era diputado, participé en el debate más general como diputado, pero también en la Comisión de Hacienda. La Comisión de Hacienda de la Cámara [de Diputados] fue la que bloqueó este proyecto e impidió que saliera y lo mandó a Comisión Mixta para reestudiar algunos aspectos. Los aspectos que yo recuerdo y diría más importantes fueron el concepto de calidad que estaba en la Agencia de Calidad, el segundo era... que todo el sistema estaba separado eran una serie de cuestiones paralelas y no tenían ninguna coordinación entre la Superintendencia, la Agencia, la Comisión Nacional de Educación y el Ministerio. Eran cuestiones totalmente separadas unas de otras. Y la cuarta cosa es que no había ninguna singularidad para la educación pública (Entrevista Parlamentario Concertación 04).

Finalmente, la aprobación del proyecto de ley permitiría la implementación del SACE, que contaría con la creación de dos nuevas instituciones, la Agencia de Calidad y la Superintendencia, que junto al Ministerio de Educación, articularían un sistema de aseguramiento encargado de la provisión de apoyo pedagógico, así como de inspeccionar, evaluar, clasificar, fiscalizar y sancionar a las escuelas en función de su desempeño. Sin embargo, como señalan Berner y Bellei (2011, p. 74), algunos de los aspectos que incorporan, como por ejemplo, “el cierre de establecimientos públicos aun en zonas en donde no haya más oferta pública”, o las sanciones asociadas al desempeño, son problemáticas debido a las dificultades de su medición y al efecto del contexto social y académico de los establecimientos en los resultados.

⁶ El Congreso de Chile está formado por dos Cámaras, la Cámara de Diputados y el Senado de la República. Cuando surgen divergencias y/o conflictos entre ambas cámaras durante el proceso de tramitación de un proyecto de ley, la Comisión Mixta (que está formada por 5 senadores y 5 diputados), se encarga de la resolución de dichas divergencias.

En definitiva, durante la década de los 2000 se llevó a cabo un complejo proceso político que derivó en la adopción de nuevos mecanismos de RdC basados en la lógica de la NGP. Contraintuitivamente, el mercado ha precisado de la intervención estatal para regular sus múltiples fallas. Sin embargo, lejos de revertir la dinámica del mercado, mediante la introducción de nuevas herramientas de control desde el Estado, se han sofisticado los mecanismos de competencia y diferenciación de la oferta escolar. Ahora bien, estas reformas no lograron satisfacer las demandas sociales formuladas por el movimiento estudiantil, sino que luego de la promulgación de la Ley SACE el debate continuó, surgiendo nuevos actores sociales, como por ejemplo, la plataforma Alto al SIMCE (ver Pino, Oyarzún & Salinas, 2016) y el Movimiento por la Unidad Docente (ver González, 2015), que han tratado de revertir la actual arquitectura del sistema educativo.

Asimismo, un creciente número de investigaciones cualitativas han discutido los efectos de las políticas de RdC, mostrando los efectos perversos y las tensiones que generan sobre los diferentes actores escolares, y especialmente sobre los docentes, en términos tanto de sus prácticas como de sus subjetividades (Acuña et al. 2014; Assaél et al., 2014; Carrasco, 2010; Falabella, 2013, 2016a). La literatura evidencia que estas políticas tensionan a los docentes y a las escuelas, desencadenando la adopción de comportamientos estratégicos (como por ejemplo, el “teaching to the test” o la reducción del currículum) para mejorar el desempeño en las pruebas de evaluación estandarizadas. A su vez, se encuentra que estas tensiones se exageran en aquellos contextos más vulnerables (Falabella, 2016b). Por otro lado, múltiples investigaciones señalan que las políticas de RdC han contribuido a reconfigurar las subjetividades de los docentes, mediante la sofisticación de los mecanismos de control y la estandarización de los objetivos y las prácticas pedagógicas, limitando así su autonomía profesional y generando nuevas “subjetividades de sumisión” ante la RdC, que se expresan mediante sentimientos de “ansiedad”, “estrés”, “agobio” y “culpa” (Cornejo et al. 2015; Fernández, 2016; Rojas & Leyton, 2014).

Conclusiones

La trayectoria de las reformas de RdC en el caso chileno es atípica, ya que la emergencia de los mecanismos de RdC surge por la necesidad de regular el mercado escolar y sus múltiples fallas, sustituyendo la “mano invisible” del mercado por la “mano visible” del Estado Evaluador. El presente artículo ha examinado los antecedentes de estas reformas, y cómo los discursos y las

soluciones políticas que circulaban a nivel global y nacional, fueron tejiendo y consolidando la idea en el campo educativo chileno de que se requería más Estado para regular y sofisticar el funcionamiento del mercado.

El artículo muestra cómo la confluencia de la crisis educativa asociada al estancamiento de los resultados en las evaluaciones estandarizadas como SIMCE y PISA, junto con las protestas protagonizadas por el movimiento estudiantil de 2006, facilitó la apertura de una “ventana de oportunidades” políticas (Kingdon, 1995) que fue aprovechada estratégicamente por los *policy-makers* para consolidar las medidas de RdC y el modelo de Estado evaluador.

A este respecto, la protesta social liderada por el movimiento estudiantil de 2006 derivó en la conformación de un Consejo Asesor Presidencial promovido por el Gobierno de la Concertación como vía para institucionalizar el conflicto social. En el Consejo Asesor Presidencial se formularon una amplia gama de propuestas para la reforma. Sin embargo, posteriormente, el Gobierno crearía una comisión técnico-política con la que se retornaría a la “política de los consensos” imperante en los sucesivos gobiernos de la Concertación, y que se definía por la negación de debates de corte estructural (Alarcón-Leiva, Johnston y Frites-Camilla, 2014). La negación de estos debates permitió mantener la armonía política del bipartidismo, pero a su vez, supuso un creciente divorcio con la ciudadanía, que reclamaba reformas más profundas.

En referencia a la selección de la política, los resultados subrayan la importancia del carácter polisémico de la RdC. Según Verger y Normand (2015), la amplia diseminación y aceptación de algunas tecnologías políticas como la RdC se debe a que actúan como un “significante flotante”, ya que pueden adaptarse muy fácilmente a los discursos de tradiciones ideológicas diferentes, como el conservadurismo, el neoliberalismo, el social-liberalismo o la socialdemocracia. A pesar de las diferencias programáticas entre el Gobierno de centro-izquierda y la derecha, existía un diagnóstico compartido y un amplio consenso en que la RdC, bajo una lógica de NGP, era la solución política más deseable. Por una parte, la RdC permite atraer y conjugar los valores enraizados en la política tradicional de la derecha, que se adhiere a una lógica de mercado y ve en la RdC un mecanismo clave para fortalecer el “derecho a la libertad de elección” y mejorar la información de los consumidores. Por su parte, el centro-izquierda, que reivindica un mayor rol del Estado en la gobernanza del sistema educativo, percibe la RdC como una herramienta

fundamental para guiar y controlar las escuelas a la distancia y promover la equidad del sistema educativo.

En efecto, la política de RdC se erigió en Chile como una propuesta factible desde el punto de vista técnico y económico, y a su vez permitía alinear entre sí las preferencias ideológicas de los *policy-makers* de las principales coaliciones políticas. Además, la racionalidad política que abogaba por la RdC también se alineaba con las ideas que circulaban a nivel global, y concretamente con las propuestas diseminadas por los organismos internacionales en el contexto del Post-Consenso de Washington.

Sin embargo, a pesar del consenso entre la derecha y el centro-izquierda sobre la necesidad de fortalecer el Estado evaluador y la política de RdC, durante la fase de promulgación e institucionalización de la política también hubo divergencias y tensiones entre ambas coaliciones, entre diferentes facciones de la coalición de centro-izquierda, y entre los discursos hegemónicos en el campo político y las reivindicaciones de los actores sociales. Los principales debates pivotaron alrededor de cuestiones como la regulación del sector privado y la prohibición (o no) de lucrar y de seleccionar al alumnado, el concepto de calidad, la problemática de la inequidad educativa y el rol preferente del Estado frente a la educación pública.

La política es “un proceso, algo en desarrollo, interaccional e inestable” (Ball, 2008, p. 7). En el caso de las reformas estudiadas, a pesar de los consensos existentes entre el Gobierno y la oposición, los principales debates que se censuraron en la arena política durante el ciclo de reformas analizado (entre los cuales destaca el rol de la educación pública y la eliminación del voucher, entre otros) siguen presentes. Asimismo, en los últimos años han surgido nuevos actores sociales en el campo educativo chileno (integrados por docentes, académicos y estudiantes) y, con ellos, también han emergido nuevas narrativas y prácticas de acción colectiva, que han puesto en cuestión las políticas de RdC y estandarización. Por lo tanto, nos encontramos ante un proceso dinámico, con resistencias y pequeños cambios desde la política (por ejemplo, se redujeron el número de pruebas SIMCE y no se continuó confeccionando “rankings de escuelas”), aunque de momento, en general la lógica de RdC persiste.

En definitiva, este artículo subraya la complejidad de los procesos de adopción de políticas y los múltiples factores, de ámbito tanto nacional como internacional, que convergen en la producción de la política educativa. En el artículo, hemos puesto de relieve que los procesos de adopción de reformas educativas son el resultado de relaciones dinámicas entre actores, discursos y estrategias, y de procesos de negociación complejos cuyos resultados son contingentes a los contextos políticos e históricos en los que se llevan a cabo.

Agradecimientos

Este estudio fue financiado por el proyecto REFORMED – “Reforming Schools Globally: A Multi-Scalar Analysis of Autonomy and Accountability Policies in Education” (ERC Starting Grant, ref. 680172), y por el proyecto Fondecyt n°11140302, Gobierno de Chile. Asimismo, la investigación también ha sido apoyada por una beca FPU del Ministerio de Educación, Cultura y Deporte español (ref. FPU13/01090).

Referencias

- Acuña, F., Assaél, J., Contreras, P., & Peralta, B. (2014). La traducción de los discursos de la política educativa en la cotidianeidad de dos escuelas municipales chilenas: La metáfora médica como vía de análisis. *Psicoperspectivas*, 13(1), 46-55. <https://doi.org/10.5027/PSICOPERSPECTIVAS-VOL13ISSU1-FULLTEXT-363>
- Alarcón-Leiva, J., Johnston, E., & Frites-Camilla, C. (2014). Consenso político y pacto educativo: pospolítica y educación en Chile (1990-2012). *Universum (Talca)*, 29(2), 37-48. <https://doi.org/10.4067/S0718-23762014000200004>
- Assaél, J., Acuña, F., Contreras, P., & Corbalán, F. (2014). Transformaciones en la cultura escolar en el marco de la implementación de políticas de accountability en Chile: Un estudio etnográfico en dos escuelas clasificadas en recuperación. *Estudios pedagógicos (Valdivia)*, 40(2), 07-26. <https://doi.org/10.4067/S0718-07052014000300001>
- Ball, S. J. (1998). Big policies/small world: An introduction to international perspectives in education policy. *Comparative education*, 34(2), 119-130. <https://doi.org/10.1080/03050069828225>
- Ball, S. J. (2008). *The education debate*. Bristol: Policy Press.

- Ball, S. J., & Youdell, D. (2008). *Hidden privatisation in public education*. Brussels: Education International.
- Banco Mundial. (2007). *El diseño institucional de un sistema efectivo de aseguramiento de la calidad de la educación en Chile*. Washington
- Bellei, C. (2015). *El gran experimento: Mercado y privatización de la educación chilena*. Santiago de Chile: LOM Ediciones.
- Bellei, C., & Cabalin, C. (2013). Chilean Student Movements: Sustained Struggle to Transform a Market-Oriented Educational System. *Current Issues in Comparative Education*, 15(2), 108-123.
- Bellei, C. & Vanni, X. (2015). The Evolution of Educational Policy in Chile 1980-2014. En S. Schwartzman (Ed.), *Education in South America*. London: Bloomsbury.
- Bloque Social. (2006). *En Chile: Diagnóstico y propuestas*. Santiago de Chile.
- Bovens, M. (2007). Analysing and assessing accountability: a conceptual framework. *European Law Journal*, (4), 447-468. <https://doi.org/10.1111/j.1468-0386.2007.00378.x>
- Berner, H., & Bellei, C. (2011). ¿Revolución o reforma? Anuncios, medidas y compromisos a la espera de la reforma educacional. *Política. Revista de Ciencia Política*, 49(2), 67-96. <https://doi.org/10.5354/0716-1077.2011.17275>
- Beyer, H. (2001). Falencias institucionales en educación: Reflexiones a propósito de los resultados del TIMSS. *Estudios Públicos*, 82, 5-33.
- Burton, G. (2012). Hegemony and frustration: education policy making in Chile under the Concertación, 1990–2010. *Latin American Perspectives*, 30(10), pp. 1-19
- Campbell, J. L. (2004). *Institutional Change and Globalization*. Princeton University Press.
- Carrasco, A. (2010). *A case-study of the Chilean policy agenda for disadvantaged primary schools: meeting their challenges?* (Unpublished doctoral dissertation). University of Cambridge.
- Consejo Asesor Presidencial para la Calidad de la Educación. (2006). *Informe Final*. Santiago: Consejo Asesor Presidencial para la Calidad de la Educación.
- Cornejo, R., Albornoz, N., Castañeda, L., Palacios, D., Etcheberrigaray, G., Fernández, R., & Lagos, J. I. (2015). Las prescripciones del trabajo docente en el nuevo marco regulatorio de políticas educativas en Chile. *Psicoperspectivas*, 14(2), 72-83.
- Cox, C. (2012). Política y políticas educacionales en Chile 1990-2010. *Revista Uruguaya de Ciencia Política*, 21(1), 13-43.

- Donoso, S. (2013). Dynamics of change in Chile: Explaining the emergence of the 2006 Pingüino movement. *Journal of Latin American Studies*, 45(01), 1-29. <https://doi.org/10.1017/S0022216X12001228>
- Fairclough, N. (1995). *Critical discourse analysis. The critical study of language*. London: Longman.
- Falabella, A. (2013). *Accountability policy effects within school markets. A study in three Chilean municipalities* (Unpublished doctoral dissertation). Institute of Education. University of London.
- Falabella, A. (2014). The Performing School: Effects of Market & Accountability Policies. *Education Policy Analysis Archives*, 22(51). <https://doi.org/10.14507/epaa.v22n70.2014>
- Falabella, A. (2015). El mercado escolar en Chile y el surgimiento de la Nueva Gestión Pública: El tejido de la política entre la dictadura neoliberal y los gobiernos de la centroizquierda (1979 a 2009). *Educação & Sociedade*, 36(132), 699-722. <https://doi.org/10.1590/ES0101-73302015152420>
- Falabella, A. (2016a). ¿Qué aseguran las políticas de aseguramiento de la calidad? Un estudio de casos en distintos contextos escolares. *Estudios pedagógicos*, 41(1), 107-126. ISSN: 0718-0705.
- Falabella, A. (2016b). Do national test scores and quality labels trigger school self-assessment and accountability? A critical analysis in the Chilean context. *British Journal of Sociology of Education*. 37(5), 743-760. <https://doi.org/10.1080/01425692.2014.976698>
- Falabella, A. (en prensa). La seducción por el Estado supervigilante en la educación escolar chilena: Un recorrido de 1979 a 2011. En C. Ruiz, F. Herrera, & L. Reyes, *Educación Pública y Privatización*. Santiago: Editorial Universitaria.
- Fernández, T. (2007). *Distribución del conocimiento escolar: clases sociales, escuelas y sistemas educativos en América Latina*, México: El Colegio de México.
- Fernández, R., Albornoz, N., Cornejo, R., & Etcheberrigaray, G. (2016). Los discursos sobre autonomía del trabajo docente en el nuevo marco regulatorio educativo chileno. *Currículo sem Fronteiras*, 16(2), 283-302.
- Fontaine, L., & Eyzaguirre, B. (2001). Una estructura que presione a las escuelas a hacerlo bien. En: H. Beyer, R. Vergara (Eds.). *¿Qué hacer ahora?* (pp. 65-111). Santiago: CEP.
- García-Huidobro, J., & Bellei, C. (2010). Debate sobre la calidad de la educación: ¿tropiezo u oportunidad?. *Revista Mensaje*, 59(589), 202-207.

- González, E. (2015). *Arriba profes de Chile: De la precarización neoliberal a la reorganización docente*. Santiago, Chile: LOM.
- Gunter, H. M., Grimaldi, E., Hall, D., & Serpieri, R. (2016). *New Public Management and the Reform of Education: European Lessons for Policy and Practice*. London: Routledge.
- Haas, P. M. (2004). When does power listen to truth? A constructivist approach to the policy process? *Journal of European Public Policy*, 11(4), 569–592. <https://doi.org/10.1080/1350176042000248034>
- Hall, P. (1993). Policy Paradigms, Social Learning and the State. The Case of Economic Policymaking in Britain. *Comparative Politics*, 25(3), 275–296. <https://doi.org/10.2307/422246>
- Hatch, T. (2013). Beneath the surface of accountability: Answerability, responsibility and capacity-building in recent education reforms in Norway. *Journal of Educational Change*, 14(2), 113-138. <https://doi.org/10.1007/s10833-012-9206-1>
- Hay, C. (2002). *Political Analysis. A critical introduction*. New York: Palgrave. <https://doi.org/10.1007/978-0-230-62911-0>
- Jessop, B. (2010). Cultural political economy and critical policy studies. *Critical Policy Studies*, 3(3-4), 336-356, <https://doi.org/10.1080/19460171003619741>
- Kingdon, J. W. (1995). *Agendas, alternatives, and public policies*. New York: HarperCollins College Publishers.
- Larroulet, C. & Montt, P. (2010). Políticas educativas de largo plazo y acuerdo amplio en educación: El caso chileno. En G. Elacqua, S. Martinic (Eds.). *¿Fin de Ciclo? Cambios en la Gobernanza del Sistema Escolar*, (pp. 19-54). Santiago: PUC, UNESCO.
- Maroy, C. (2009). Convergences and hybridization of educational policies around ‘postbureaucratic’ models of regulation. *Compare*, 39(1), 71-84. <https://doi.org/10.1080/03057920801903472>
- Mineduc (2016). *Plan de Evaluaciones Nacionales e Internacionales 2016-2020*. Santiago: Ministerio de Educación, Gobierno de Chile.
- Mizala, A., & Schneider, B. R. (2014). Negotiating education reform: Teacher evaluations and incentives in Chile (1990–2010). *Governance*, 27(1), 87-109. <https://doi.org/10.1111/gove.12020>

- Montt, P., Elacqua, G., Pacheco, P., González, P., & Raczynski, D. (2006). *Hacia un Sistema escolar descentralizado, sólido y fuerte. El diseño y las capacidades hacen la diferencia*. Santiago: MINEDUC.
- Newman, J. (2005) *Remaking governance. Peoples, politics and the public sphere*. Bristol: The Policy Press.
- OCDE. (2004). *Reviews of National Policies for Education*. Chile. París: OECD. <https://doi.org/10.1787/9789264106352-en>
- OCDE. (2011). PISA in Focus 9: School autonomy and accountability: Are they related to student performance?.París: OECD. <http://www.oecd.org/pisa/pisaproducts/pisainfocus/48910490.pdf>
- Parcerisa, L., & Verger, A. (2016). Rendición de cuentas y política educativa: Una revisión de la evidencia internacional y futuros retos para la investigación. *Profesorado, revista de currículum y formación del profesorado*, 20(3), 15-51. <https://recyt.fecyt.es/index.php/profesorado/article/view/54589/33220>
- Pino, M., Oyarzún, G., & Salinas, I. (2016). Crítica a la estandarización para la rendición de cuentas: narrativas de resistencia para transformar el sistema de evaluación en Chile, *Cadernos Cedes*, 36(100), 337-354. <https://doi.org/10.1590/cc0101-32622016171362>
- Rojas, M. T., & Leyton, D. (2014). La nueva subjetividad docente: Construcción de subjetividades docentes en los inicios de la implementación de la Subvención Escolar Preferencial en Chile. *Estudios pedagógicos (Valdivia)*, 40(especial), 205-221. <https://doi.org/10.4067/S0718-07052014000200012>
- Rose, N., & Miller, P. (1992). Political power beyond the state: problematics of government. *British Journal of Sociology*, 43(2), 173-205. <https://doi.org/10.2307/591464>
- Santa Cruz, E. (2016). *Mediatización de las políticas educativas en Chile: El discurso de los diarios La Tercera y El Mercurio sobre la Ley General de Educación (2006-2009)*. Tesis Doctoral, Universidad de Granada.
- Schriewer, J. (1990). The Method of Comparison and the Need for Externalization: Methodological Criteria and Sociological Concepts. En J. Schriewer & B. Holmes (Eds.) *Theories and Methods in Comparative Education* (pp. 25-83). Frankfurt am Main: Peter Lang.
- Traverso, A. (2004). *Las nuevas reformas que la Reforma engendra*. En foco, 26. Documento Expansiva. Santiago.

- Valenzuela, J. P., Bellei, C., & De Los Ríos, D. (2010). Segregación escolar en Chile. *En: S. Martinic & G. Elacqua, (Eds.) ¿Fin del ciclo? Cambios en la gobernanza del sistema educativo* (pp. 209-229). Santiago de Chile: UNESCO and Pontificia Universidad Católica de Chile.
- Valenzuela, J. P., Bellei, C., & De Los Ríos (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy*, 29(2), 217-241. <https://doi.org/10.1080/02680939.2013.806995>
- Vera, S. (2011). Cronología del Conflicto: El Movimiento Estudiantil en Chile, 2011. *En: Anuario del Conflicto Social 2011*, Barcelona: Observatorio del Conflicto Social. pp. 252-261.
- Verger, A. (2014). Why Do Policy-Makers Adopt Global Education Policies? Toward a Research Framework on the Varying Role of Ideas in Education Reform. *Current Issues in Comparative Education*, 16(2), 14-29.
- Verger, A. (2016). The global diffusion of education privatization: unpacking and theorizing policy adoption. *En K. Mundy, A. Green, B. Lingard, & A. Verger, Handbook of Global Education Policy* (pp. 64-80). John Wiley & Sons. <https://doi.org/10.1002/9781118468005.ch3>
- Verger, A., BonaL, X., & Zancajo, A. (2016). What Are the Role and Impact of Public-Private Partnerships in Education? A Realist Evaluation of the Chilean Education Quasi-Market. *Comparative Education Review*, 60(2), 223-248. <https://doi.org/10.1086/685557>
- Verger, A., Curran, M., & Parcerisa, L. (2015). La trayectoria de una reforma educativa global: El caso de la Nueva Gestión Pública en el sistema educativo catalán. *Educação & Sociedade*, 36(132), 675-697. <https://doi.org/10.1590/ES0101-73302015152619>
- Verger, A., Fontdevila, C., & Zancajo, A. (2016). *The Privatization of Education: A Political Economy of Global Education Reform*. New York: Teachers College Press.
- Verger, A., & Normand, R. (2015). New public management and education: theoretical and conceptual elements for the study of a global education reform model. *Educação & Sociedade*, 36(132), 599-622. <https://doi.org/10.1590/ES0101-73302015152799>
- Whitty, G. (2012). Policy Tourism and Policy Borrowing in Education: A Trans-Atlantic Case Study. *En G. Steiner-Khamsi & F. Waldow (Eds.) World Yearbook of Education 2012: Policy Borrowing and Lending in Education*, (pp. 334-370). London: Routledge.
- World Bank. (2015). What Matters Most for School Autonomy and Accountability: A Framework Paper. SABER Working Paper Series, 9. Recuperado de:

https://olc.worldbank.org/sites/default/files/Whatmatters_SAA_framework_paper_Demas_2014_4.pdf

Zancajo, A. (2017). *Las escuelas ante el mercado. Análisis de las respuestas de la oferta escolar en un entorno de competencia: el caso de Chile*. Unpublished PhD dissertation, Universitat Autònoma de Barcelona.

POLICY ENACTMENT

Publication C: Parcerisa, L., & Verger, A. (2016). Rendición de cuentas y política educativa: Una revisión de la evidencia internacional y futuros retos para la investigación [*Accountability and education policy: A review of international evidence and future research challenges*]. *Profesorado. Revista de Currículum y Formación de Profesorado*, 20(3), 15-51.
<https://recyt.fecyt.es/index.php/profesorado/article/view/54589/33220>

CHAPTER 3

Rendición de cuentas y política educativa: una revisión de la evidencia internacional y futuros retos para la investigación

Introducción

La rendición de cuentas (RdC) está adquiriendo cada vez más centralidad en la agenda educativa global. A pesar de que la RdC no es una nueva política en el campo educativo, las formas, el rol y los objetivos principales de la RdC han cambiado sustancialmente en los últimos años. Tradicionalmente, la RdC de carácter político y/o legal ha sido implementada con el objetivo de aumentar el control democrático y ciudadano de la educación. Sin embargo, últimamente, la RdC ha adquirido un carácter más bien gerencial (en el sentido que se basa en la obtención y evaluación de resultados escolares) y, además, ha pasado a ser concebida como una de las soluciones más importantes a los múltiples problemas y retos de los sistemas educativos contemporáneos. Actualmente, la RdC en educación es considerada como una fuente de eficiencia (es decir, una forma de promover una mejor alineación entre las aspiraciones gubernamentales y los propósitos de la escuela), de excelencia académica (una herramienta para mejorar el desempeño por parte de los centros, los profesores y los estudiantes), y de equidad (una manera de garantizar que todos los estudiantes alcancen un nivel mínimo de competencias básicas).

Las organizaciones internacionales más influyentes en el campo educativo han alimentado tan altas expectativas respecto a la RdC. En este sentido, las políticas de RdC son absolutamente centrales en las recomendaciones de los informes sobre educación elaborados por la Organización para la Cooperación y el Desarrollo Económico (OCDE), incluido el influyente *Programa para la Evaluación Internacional de Estudiantes* (PISA). De acuerdo con un reciente estudio, 29 representantes de los países de la OCDE (de 37) admitieron que las recomendaciones de PISA/OCDE sobre RdC han influido en los procesos de formulación de políticas sobre esta temática a nivel nacional (Breakspear, 2012). Para el Banco Mundial, la RdC a nivel escolar es una condición necesaria para mejorar la calidad de los maestros y el aprendizaje, ya que permite alinear los incentivos de los diferentes agentes educativos (Banco Mundial, 2015). Por su lado,

la UNESCO da mucha importancia a la RdC en su *Marco de Acción de Educación 2030*, y dedica su influyente *Informe de Seguimiento de la Educación en el Mundo* del año 2017 a este tema. Esta organización internacional considera que la introducción de mecanismos de RdC en los sistemas nacionales de educación es una condición necesaria para que los países cumplan con los objetivos de desarrollo sostenible en educación (UNESCO, 2015).

En términos de políticas, la RdC es un concepto polisémico que cubre una amplia gama de opciones y de modelos (incluyendo los modelos legal, político, burocrático y de mercado). No obstante, el modelo de RdC que actualmente está ganando más peso a nivel internacional está fundamentado en la evaluación administrativa y externa de los resultados de aprendizaje, y se centra en las escuelas y/o en los maestros como los actores que rinden cuentas. Este tipo de RdC se conoce generalmente como RdC gerencial o basada en pruebas (*testing-based accountability*) (Kamens y Benavot, 2011; Tobin, et al., 2015).

A medida que un mayor número de países adoptan sistemas de RdC - y, en particular, sistemas de RdC basados en pruebas - dichos sistemas se convierten en el foco de más investigaciones académicas. No obstante, la investigación disponible sobre la temática todavía es insuficiente para comprender en qué circunstancias y a través de qué mecanismos las políticas de RdC podrían generar los resultados esperados. De hecho, como demuestra este artículo, la evidencia sobre los efectos de las políticas de RdC sobre los resultados del aprendizaje, pero también sobre otros aspectos importantes de la educación, tales como la autonomía de los maestros, la organización de las escuelas o los procesos educativos aún no es concluyente y refleja efectos diferentes (e incluso contradictorios) en distintos lugares.

En este artículo se revisa la evidencia internacional sobre los efectos de las políticas de RdC en la educación. En concreto, sobre la base de la evidencia existente, nuestro objetivo es mostrar cómo los diferentes componentes de los sistemas de RdC (es decir, el diseño de dichos sistemas, los procesos de recontextualización [o puesta en práctica] de las políticas, las contingencias contextuales y su impacto) interactúan de forma compleja, y en qué sentido. Las preguntas que guían nuestra investigación son: 1) ¿Cuáles son los principales efectos de las políticas de rendición de cuentas en educación?; 2) ¿Cómo se manifiestan estos efectos en diferentes

dimensiones educativas y cuáles son las principales áreas de acuerdo y desacuerdo?; 3) ¿Bajo qué circunstancias tienen las políticas de RdC los efectos esperados?

Con este objetivo en mente, el artículo está organizado de la siguiente manera. En la primera sección, se presentan los principales modelos de RdC en educación, y se explica por qué la RdC basada en pruebas se está convirtiendo en el modelo predominante a nivel internacional. En la segunda sección del artículo se muestra cómo la RdC basada en pruebas puede cristalizar en diseños muy diferentes de política, y por qué las decisiones sobre cuestiones de diseño son tan decisivas a la hora de entender la trayectoria de las reformas de RdC. En tercer lugar, se reflexiona sobre las diferentes dimensiones de impacto de las políticas de RdC en educación, incluyendo resultados inesperados o no deseados a nivel de escuela y de enseñanza. En cuarto lugar, nos centramos en las dinámicas y los procesos de recontextualización de las políticas de RdC - incluyendo cómo los maestros reciben, resisten y/o transforman la RdC a través de sus prácticas diarias. En quinto lugar, revisamos cómo y en qué medida los contextos institucionales y socioeconómicos en los que se aplican las políticas de RdC, también moldean la trayectoria e impacto de dichas políticas. En la sección final, en base a nuestra revisión, extraemos premisas clave y direcciones para futuras investigaciones.

Metodológicamente, el capítulo se basa en el enfoque de la revisión de alcance (o *scoping review*) de la literatura (Alegre, 2015; O’Flaherty y Phillips, 2015). Este enfoque de revisión de la literatura, en contraste con la revisión sistemática de la literatura, no parte de una pregunta de investigación muy específica ni intenta probar una teoría concreta. La *scoping review* está particularmente bien equipada para identificar las principales áreas de acuerdo y desacuerdo dentro de un determinado campo de trabajo (en nuestro caso la RdC en educación), así como las principales lagunas en el corpus actual de la literatura en este campo, en un período relativamente corto de tiempo (Mays, Roberts y Popay, 2001; Arksey y O’Malley, 2005). En este sentido, es importante subrayar que la amplitud, la profundidad y exhaustividad de la *scoping review* dependerá del propósito y/o los objetivos de la revisión. Esta metodología está compuesta por cinco fases: 1) identificación de los objetivos y de las preguntas que guían la revisión (las cuales hemos mencionado anteriormente); 2) identificación de la literatura sobre la temática; 3) selección de las investigaciones más relevantes para la revisión; 4) catalogación e interpretación de los datos; 5) ordenación, resumen y comunicación de los principales hallazgos

de la revisión (Arksey y O'Malley, 2005). La búsqueda bibliográfica se realizó a partir de las bases de datos de SCOPUS y Web of Science (WoS) y se articuló en dos niveles combinando diferentes palabras clave (véase apéndice 1): un primer nivel, relativo a la adopción y aplicación de las políticas de RdC, y el segundo nivel, relacionado con los efectos de las políticas.

La búsqueda excluyó la educación superior y se limitó al área de ciencias sociales. El límite temporal se fijó entre los años 1995 y 2014. En total, 277 documentos fueron identificados en la primera búsqueda. Seguidamente, se aplicó otro filtro relacionado durante la lectura de los resúmenes. En este caso, se descartaron aquellos estudios que no eran relevantes para el objeto de la investigación, o bien no aportaban evidencia empírica. Durante el proceso iterativo de selección se llevaron a cabo búsquedas complementarias en otras bases de datos, como por ejemplo Google Scholar. Finalmente, 150 documentos fueron revisados, aunque no todos han sido utilizados para la elaboración de éste artículo. La información de los artículos se recolectó en fichas de lectura que incluían los siguientes ítems (adaptados a partir de Arksey y O'Malley, 2005): 1) datos de identificación: autores, año, referencia de la publicación y localización del estudio; 2) tipo de política; 3) objetivos y/o hipótesis de la investigación; 4) metodología; y 5) resultados: proceso de adopción y efectos. Por último, se ordenaron los resultados en base a las diferentes dimensiones identificadas y se llevó a cabo la síntesis de la literatura.

La emergencia de la rendición de cuentas basada en pruebas estandarizadas

En términos generales, la RdC remite a los procesos, mecanismos y/o instrumentos que hacen que las organizaciones (pero también los individuos que integran las organizaciones) cumplan con sus obligaciones y se vuelvan más sensibles a sus públicos específicos (Bovens 2007, Hatch 2013). Tanto la obligatoriedad como la capacidad de condicionar la respuesta de los sujetos son características inherentes a los sistemas de RdC. De acuerdo con Bovens (2007):

Accountability is a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences (Bovens 2007, p.450).

Sin embargo, a pesar de estas características muy generales, la RdC es un concepto polisémico que cubre una amplia gama de enfoques y tipologías de políticas. Las tipologías más tradicionales

de RdC son la política, la legal y la burocrática. No obstante, otros modelos de RdC presentes en determinados países o regiones son el modelo profesional (según el cual los maestros son responsables de respetar las normas profesionales y deben rendir cuentas a sus iguales –o a las organizaciones profesionales – Anderson, 2005); el modelo de mercado (que se basa en rendir cuentas a los consumidores y se promueve a través de las políticas a favor de la libertad de elección de centro – West, et al., 2011); o el modelo de RdC participativa o social, que significa que las escuelas y sus profesores son responsables ante las familias a través de la voz y el diálogo dentro de los órganos de gobierno de la escuela (West, et al., 2011). En los últimos años múltiples investigadores han desarrollado taxonomías sobre los diferentes modelos de RdC. En la Tabla 2 se presentan algunas de las principales taxonomías de la RdC en el campo educativo. Esta tabla de síntesis incorpora las principales categorías que se encuentran en la literatura (por ejemplo, la categoría de mercado, la profesional o la política), y al mismo tiempo, muestra otras categorías que a pesar de citarse bajo diferentes etiquetas, tienen un significado muy parecido (por ejemplo, en el caso de la burocrática y la jerárquica, o la contractual y la legal). La taxonomía de West et al. (2011) es la que descompone más las distintas modalidades de RdC, incorporando la modalidad en red y la participativa. Sin embargo, la taxonomía elaborada por Maroy y Voisin (2013) resulta innovadora respecto las otras, en el sentido que es la única que no se orienta en base a la naturaleza del foro, sino en la naturaleza de la obligación e incorpora la ontología de la política.

Tabla 2.

Taxonomías de la rendición de cuentas en educación

Leithwood y Earl (2000)	Darling-Hammod (2004)	West et al. (2011)	Maroy y Voisin (2013)
<ul style="list-style-type: none"> • Profesional • Gerencial • Descentralización • Mercado 	<ul style="list-style-type: none"> • Política • Legal • Burocrática • Profesional • Mercado 	<ul style="list-style-type: none"> • Professional • Jerárquica • Mercado • Contractual • En red • Participativa 	<ul style="list-style-type: none"> • <i>Strong accountability</i> • Neo-burocrática • Reflexiva • <i>Soft accountability</i>

Fuente: autores

No obstante, de todos los modelos de RdC emergentes, el que tienen un alcance más global es el gerencial. La característica más distintiva de la RdC gerencial es su claro enfoque en los resultados (generalmente, resultados de aprendizaje), en lugar de poner énfasis en los *inputs* y los procesos. La aplicación de este enfoque de la RdC implica la generación de datos a través de instrumentos de evaluación estandarizada a gran escala. Es por ello que este enfoque de la RdC es también conocido como la RdC performativa (Ranson, 2003), RdC orientada a los resultados (Anderson, 2005) o RdC basada en pruebas (Hamilton, et al., 2007).

El modelo de RdC gerencial o basado en pruebas ha ganado popularidad y atracción internacional por diferentes razones. Para empezar, la mejora de los resultados de aprendizaje se ha convertido en el foco principal de las agendas de reforma educativa, tanto en los países del norte como del sur. De acuerdo a ello, la medición de los resultados de aprendizaje es una herramienta cada vez más necesaria para comprobar si las reformas logran los resultados esperados y para la promoción de la mejora escolar. En este terreno, la reforma *No Child Left Behind* adoptada en los EE.UU. en el año 2001 se ha convertido en una de las iniciativas más emblemáticas a nivel internacional. No obstante, más allá de la legislación nacional, las pruebas estandarizadas internacionales como PISA se han convertido en un *factor* clave de las reformas de RdC basadas en pruebas. PISA no sólo ha contribuido a introducir presión competitiva a los países para mejorar los resultados de aprendizaje, sino que también se ha convertido en un referente fundamental a la hora de transferir la tecnología necesaria para adoptar pruebas de aprendizaje a nivel nacional. El surgimiento de una industria de pruebas educativas (*school testing industry*) a nivel mundial también ha presionado en una dirección similar (véase Hogan, et al., 2016).

Cuando se trata de entender el surgimiento de la RdC basada en pruebas también tenemos que hacer referencia a la emergencia de paradigmas de reforma del sector público más amplios, como la nueva gestión pública (NGP). La NGP ha contribuido a que los decisores de políticas estén más inclinados a promover estilos de gobernanza en los servicios públicos gerencialistas y orientados hacia el logro de resultados tangibles y medibles (Kalimullah, Ashraf, y Ashaduzzaman, 2012). La NGP se asocia también a corrientes de reforma educativa, tales como

el movimiento de las competencias básicas comunes (*common core standards movement*), que requieren también de la implementación de sistemas de RdC basados en pruebas.

Sistemas de rendición de cuentas basados en pruebas de aprendizaje: Múltiples opciones de diseño

Un mismo enfoque de RdC puede adoptar formas muy diferentes en función de quién se espera que rinda cuentas, a quién se debe rendir cuentas, sobre qué se rinde cuentas, y cuáles son las consecuencias (Leithwood y Earl, 2000). En esta sección, se muestra cómo estas variables de diseño pueden interactuar y se pueden combinar de manera diferente en el marco del modelo de RdC basado en pruebas de aprendizaje.

En relación a *quién* se espera que rinda cuentas en los regímenes de RdC basados en pruebas, el *locus* se podría ubicar en el profesor individual, en grupos de profesores (por ejemplo, profesores de un mismo curso o nivel), en el director, en toda la escuela, o en redes de escuelas. El *quién* es una variable importante en el sentido de que diversos estudios muestran que, cuánto más individual sea el enfoque, más se socava la cooperación y el trabajo colegiado entre los maestros (Jones y Egley, 2004).

Los sistemas de RdC basados en pruebas de aprendizaje requieren de la presencia de una agencia de evaluación externa (que puede ser más o menos independiente del gobierno), aunque la inspección, las familias o las organizaciones profesionales pueden tener también un rol importante. Por lo que respecta a cómo se ejecuta la RdC, una variable importante es si los resultados de las evaluaciones de los centros se hacen públicos o no, y cómo (por ejemplo, si se tiene en cuenta la composición socioeconómica de la escuela o no a la hora de publicar los resultados). No obstante, la publicación de rankings escolares es una medida de RdC controvertida que, según algunos autores podría contribuir a mejorar el rendimiento académico (Boarini y Lüdemann, 2009), pero según otros podría socavar dinámicas de cooperación entre escuelas, o contribuir a la estigmatización de algunos centros (Jones y Egley, 2004).

En los regímenes de RdC basados en pruebas de aprendizaje, los resultados de los estudiantes son, casi por definición, el *objeto central* del sistema de RdC. Sin embargo, los resultados de

aprendizaje se podrían combinar con la evaluación de otros componentes, como otro tipo de resultados educativos (como el nivel de abandono escolar o las tasas de graduación), los recursos y las instalaciones de la escuela, y/o los procesos educativos (a través, por ejemplo, de observaciones en el aula, revisión de portafolios docentes, etc.). En consecuencia, los sistemas de RdC podrían ser más o menos amplios por lo que respecta a los componentes educacionales que cubren. Los diseños de RdC más comprehensivos, debido al hecho de que contemplan una gama más amplia de indicadores, podrían prevenir que las escuelas se centren en tipos específicos de resultados o pongan un excesivo énfasis en enseñar partes concretas del currículum (a costa de otras).

Por último, las *consecuencias* de las evaluaciones representan una dimensión de gran relevancia en los sistemas de RdC. En este ámbito, las opciones de diseño incluyen si la RdC es de altas consecuencias (o de alto riesgo) o no, y cuál es la naturaleza de las consecuencias: incentivos económicos, programas de formación docente, sanciones (como la intervención de la escuela, el cierre o su conversión en una escuela *chárter*), etc. Al mismo tiempo, los incentivos económicos se pueden dar a nivel de escuela o a nivel individual (en forma de bonos salariales). El importe del incentivo -si es más simbólico o más sustantivo- también condiciona el nivel de impacto en el comportamiento de la escuela o de los docentes (Escardíbul, 2015). En general, las evaluaciones de altas consecuencias ponen más presión sobre las escuelas que otras formas de evaluación, pero esto no es necesariamente algo positivo. Una presión excesiva y sin suficiente apoyo pedagógico podría promover un tipo de comportamiento oportunista o no deseado (Ohemeng y McCall-Thomas, 2013), el cual explicaremos con más detalle en la siguiente sección. En resumen, la diversidad de opciones de política examinadas muestran que no es apropiado hablar de RdC en la educación como una política concreta. Es por ello que a la hora de realizar evaluaciones más rigurosas y precisas de los efectos de la RdC es preciso desentrañar primero las características específicas del diseño de los sistemas de RdC. En esta sección hemos dado algunos ejemplos de cómo diferentes opciones de diseño de la RdC basada en pruebas de aprendizaje podrían conllevar consecuencias distintas. A continuación, exploramos más a fondo los múltiples efectos de la RdC en ámbitos como la organización escolar, el profesionalismo docente, los procesos educativos y los resultados de aprendizaje de los estudiantes.

Los múltiples efectos de la rendición de cuentas en educación

Actualmente muchos países están adoptando sistemas de RdC basados en pruebas convencidos de que se trata de una buena forma de promover la excelencia académica y la mejora de los resultados del aprendizaje. Sin embargo, los sistemas de RdC también tienen efectos en otras dimensiones educativas. Algunos de estos efectos son deseados y esperados (es decir, son efectos contemplados en el diseño de la política), mientras que otros no son ni deseados ni esperados.

El carácter mixto de los efectos de la rendición de cuentas

En relación al impacto en los resultados de aprendizaje, la investigación existente sobre RdC en el campo educativo muestra resultados bastante contradictorios. Varios investigadores consideran que existe una relación positiva entre las políticas de RdC y el rendimiento académico de los estudiantes, aunque muchos de ellos alertan que, por diferentes razones, esta relación puede ser bastante débil. Por ejemplo, en una investigación llevada a cabo en Israel, Rosenblatt y Shimoni (2001) encuentran que existe una asociación positiva entre la participación de los profesores de educación física en los programas de RdC y el rendimiento de los estudiantes. Sin embargo, los autores reconocen que la muestra que utilizaron era pequeña y presentaba algunas limitaciones.

Del mismo modo, Chiang (2009) aplica un análisis de regresión discontinua con datos administrativos de Florida, y llega a la conclusión de que la amenaza de sanciones a las escuelas de bajo rendimiento genera una mejora en los resultados de matemáticas de los estudiantes de primaria que persisten durante los primeros años de la escuela media, aunque este autor señala también que dicha mejora no se pone de manifiesto en el ámbito de la lectura. Por su parte, Roderick et al. (2002), en base a datos administrativos de estudiantes de las Escuelas Públicas de Chicago, encuentran un efecto muy positivo en la mejora de las competencias básicas en dos grados (6° grado y 8° grado), aunque dicha mejora no se dio en los estudiantes de 3er grado.

Otras investigaciones son aún más explícitas acerca de la naturaleza mixta de los resultados encontrados. Este es el caso de Boarini y Lüdemann (2009), quienes concluyen que, aunque algunos aspectos de las políticas de RdC (por ejemplo, los exámenes nacionales estandarizados externos) pueden estar asociados con mejores resultados de los estudiantes, otros tienen efectos

pequeños sobre los resultados de los estudiantes (por ejemplo, la publicación de los resultados de las escuelas), o no tienen. Por su parte, Ladd (1999) encuentra efectos positivos y robustos de la RdC en Dallas (Texas) para los estudiantes blancos e hispanos de séptimo grado, mientras que los estudiantes afro-americanos obtienen resultados menos positivos. Reback et al. (2014) examinan los efectos de la ley *No Child Left Behind* (NCLB) y encuentran tanto efectos positivos como efectos neutros en relación con el rendimiento académico y el disfrute en el proceso de aprendizaje. Dee y Jacob (2011) también se centran en los efectos de la ley NCLB sobre la puntuación de los estudiantes en las pruebas estandarizadas nacionales, y encuentran un efecto positivo estadísticamente significativo sobre los resultados de los estudiantes de 4º grado en matemáticas, pero no encuentran evidencia de mejora en Lectura.

Por último, otro grupo de investigadores considera que las reformas de RdC tienen efectos negativos en los resultados del aprendizaje, sobre todo desde una perspectiva de la equidad. Por ejemplo, según Andersen (2008), la aplicación de una reforma basada en la gestión del rendimiento (*performance management reform*) en las escuelas danesas tuvo efectos negativos sobre el rendimiento de los estudiantes con bajo estatus socioeconómico (SES). Autores como Ryan (2004) señalan que el régimen de RdC basado en la prueba promovido por la ley NCLB en Estados Unidos crea incentivos que contrarían la mejora del rendimiento de los estudiantes con bajo SES. Del mismo modo, Powers (2003) muestra que la RdC basada en pruebas promovió la segregación escolar y el abandono de los estudiantes pobres y pertenecientes a minorías étnicas en California. Este tipo de RdC también puede crear incentivos para que los gobiernos reduzcan los estándares académicos y desalienta que los mejores docentes trabajen en escuelas con altos índices de población socialmente vulnerables o con bajo rendimiento, tema que retomamos en la sección final de este artículo.

La literatura revisada también contempla dimensiones de impacto que van más allá de los resultados de aprendizaje, aunque lo hace con menos frecuencia. Se trata del impacto de la RdC en, por un lado, la relación entre escuelas y, por el otro, en la identidad y el profesionalismo docente. De acuerdo a la literatura existente, las políticas de RdC pueden desencadenar tanto la colaboración como la competencia entre escuelas. En el marco de los sistemas RdC, el contexto escolar y el papel del director parecen mediar estratégicamente sobre si el comportamiento de las escuelas deviene más cooperativo o más competitivo. Se trata de un tema de gran relevancia

dados los beneficios potenciales que las dinámicas de cooperación pueden generar en los sistemas educativos (Sahlberg, 2010; Keddie, 2015), pero que todavía ha sido poco explorado.

En cuanto a la identidad profesional y a la autonomía profesional de los docentes, la investigación existente -por lo general de carácter etnográfico- apunta que las políticas de RdC basadas en el desempeño tienen más desventajas que beneficios. Estas políticas tienden a limitar la autonomía profesional de los docentes, a generar mayores niveles de estrés, y/o a erosionar las relaciones sociales dentro de la escuela (ver Day, 2002; Jeffrey y Woods, 1996; Maxcy, 2009). A pesar de las implicaciones obvias de estos tipos de efectos en los niveles de motivación docente, también hay investigaciones que contemplan que la RdC tiene efectos positivos en la motivación del profesorado (Finnigan, 2010).

Efectos no deseados

Son cada vez más numerosos los trabajos de investigación que reflejan los efectos no deseados de las políticas de RdC en educación, especialmente de las políticas de RdC de altas consecuencias y que promueven dinámicas competitivas. En ciertos contextos, estos tipos de RdC pueden alentar la emergencia de comportamientos oportunistas por parte de escuelas, maestros y directores. El concepto de comportamiento *oportunista* se refiere a una amplia gama de prácticas, incluyendo la selección de estudiantes (Jennings, 2010), la enseñanza orientada a la prueba o *teaching to the test* (Ohemeng y Thomas-McCall, 2013), el triaje educativo (Booher-Jennings, 2005; Perryman, et al., 2011; Reback, 2008), e incluso la realización de trampas en el momento de tomar la prueba (Jacob y Levitt, 2003). Los actores educativos adoptan este tipo de comportamientos con el fin de maximizar sus resultados en las pruebas. Estos comportamientos emergen con más fuerza entre aquellas escuelas y/o aquellos docentes sometidos a mayores niveles de presión externa para lograr mejores resultados educativos, sobre todo debido a la amenaza de sanciones que estos sistemas implican en el caso de tener bajo rendimiento.

Un tipo de comportamiento oportunista identificado por la literatura es el llamado triaje educativo que, en palabras de Ladd y Lauen (2010, p.429), se refiere a “*triaging out' students well below and well above grade level and 'triaging in' students close to grade level*”. Estos autores, en su análisis de los efectos distributivos de diferentes programas de RdC en Carolina del Norte, no encuentran evidencia de la existencia de la práctica del triaje. Sin embargo, sus resultados

contrastan con los de otro estudio realizado en Texas, el cual sí demuestra que los docentes responden a los incentivos para mejorar los resultados centrándose en el grupo de estudiantes que más se aproxima a la tasa de aprobación, a expensas de otros grupos de estudiantes (Booher-Jennings, 2005).

Ohemeng y McCall-Thomas (2013) llevaron a cabo un estudio sobre los comportamientos no deseados generados por las pruebas estandarizadas en Ontario. En esta investigación se llevaron a cabo entrevistas con los diferentes agentes educativos clave con el fin de comprender las presiones e incentivos que emanan de un sistema de pruebas estandarizadas. Muchos de los profesores entrevistados argumentaron que la presión para obtener buenos resultados en las pruebas los condujo a adoptar prácticas de *teaching to the test* y a centrarse en aquellas áreas del currículum que serían evaluadas en las pruebas (Ohemeng y Thomas McCall, 2013). Del mismo modo, Jones y Egley (2004) encontraron que una de las quejas más frecuentes de los profesores en relación con el sistema de pruebas de altas consecuencias fue la necesidad de pasar gran parte de su tiempo preparando a sus estudiantes para realizar la prueba de forma efectiva –a través de la enseñanza de contenidos similares a los incluidos en la prueba y de la realización de simulacros de pruebas. Estos hallazgos son consistentes con los resultados de otros estudios (véase Au, 2007; Hamilton, et al., 2007; Linn, 2000; Menken, 2006; Volante, 2004). Se trata sin duda de prácticas que resultan preocupantes desde el punto de vista educativo ya que conllevan una reducción del currículum y socavan la idea de la enseñanza y el aprendizaje como un proceso creativo y comprehensivo.

El proceso de adopción y de recontextualización de las políticas de rendición de cuentas

La adopción de nuevas formas de RdC en educación acostumbra a ser un proceso político que genera algún nivel de conflicto y resistencia, lo cual acaba revirtiendo en la calidad de la implementación de dichas políticas y en sus resultados finales.

Estudiar la trayectoria de las reformas educativas y cómo dicha trayectoria condiciona los resultados de las reformas no es una tarea sencilla. Según McDermott (2007), en toda reforma educativa acostumbra a haber mucha diferencia entre, por un lado, los objetivos o las aspiraciones declaradas de la reforma y, por el otro, las prácticas que se llevan a cabo en la

realidad en nombre de dicha reforma. Que una política educativa cuente con un buen diseño y sea promovida con buenas intenciones es una condición necesaria pero no suficiente para que la política sea exitosa. En general, la percepción y la adhesión de los agentes educativos hacia la política en cuestión resulta clave a la hora de comprender los éxitos y fracasos de dicha política. Esto es especialmente relevante en el caso de la adopción de sistemas de RdC ya que se trata de reformas que acostumbran a generar controversia y respuestas diferenciales por parte del profesorado, incluyendo estrategias de negociación y de resistencia tanto abiertas como encubiertas.

Según Ball et al. (2012), en la literatura sobre política educativa, el concepto de implementación de políticas es utilizado a menudo de forma restringida y lineal, ya que omite los procesos (micro)políticos que se dan en las escuelas. Estos autores sugieren que los textos y los programas de políticas no son implementados de manera mecánica, sino que son interpretados y recontextualizados de forma activa por parte de los actores educativos a nivel micro-institucional de formas a menudo poco previsibles. Es por ello que en lugar del concepto de implementación, Ball et al. (2012) sugieren utilizar el concepto de “*policy enactment*”, el cual sirve para capturar y comprender el proceso contingente, y su vez creativo, de decodificación y recodificación mediante el cual las políticas se ponen en acción en las escuelas.

Por lo tanto, para tener una visión comprehensiva y global de las reformas educativas y entender mejor sus efectos es necesario abrir la caja negra que hay entre el diseño y los resultados de las políticas, y analizar aquellos procesos de carácter político y/o semiótico mediante los que las políticas se traducen en prácticas concretas. Esto significa, en el caso que nos ocupa, analizar el proceso mediante el cual las reformas de RdC han sido adoptadas y recibidas por parte de los actores educativos encargados de aplicarlas. A continuación, exploramos el rol de los docentes en tanto que intérpretes, traductores y productores de las reformas de RdC.

El personal docente ante las políticas de rendición de cuentas

La complicidad de los diferentes actores que operan a nivel escolar (en particular los directores de centro y los docentes) es una condición necesaria para garantizar que la política sea aplicada con éxito. No obstante, para que una reforma disponga de la complicidad de los actores es

necesario que estos perciban que la reforma tendrá un impacto positivo y que la consideren significativa para su desarrollo profesional y para la mejora escolar.

En un estudio comparativo desarrollado a nivel europeo, Müller y Hernández (2010) exploraron las percepciones de los docentes sobre los sistemas de RdC. En esta investigación, aplicaron una encuesta a una muestra de 1.100 docentes de cada país (Finlandia, Suecia e Irlanda), 44 entrevistas basadas en historias de vida con un total de 22 docentes, observación y grupos de discusión. Sus resultados muestran que en Europa predomina el escepticismo y la desconfianza entre los docentes respecto a los sistemas de RdC (en particular, sobre los sistemas de RdC basados en el desempeño o en pruebas). Al mismo tiempo, su estudio señala el riesgo de que los sistemas de RdC se conviertan en un mecanismo vacío que sólo produzca una presión negativa, y que no consiga motivar a los docentes. Para evitar que la RdC se convierta en algo superfluo o en una influencia negativa para la mejora escolar, estos investigadores sugieren que los mecanismos de RdC deben orientarse a reforzar la profesionalización docente y responder a las necesidades reales de la comunidad educativa (Müller y Hernández, 2010).

De manera similar, autores como Jones y Egley (2004) exploraron las percepciones de los docentes sobre el sistema de exámenes de altas consecuencias a través de una encuesta aplicada a 708 docentes del estado de Florida, según la cual la mayoría de los docentes (79,9%) creía que las políticas de RdC implementadas en Florida no iban en la buena dirección. En base a dicho resultado, Jones y Egley (2004) concluyen que estas percepciones negativas condicionan en gran medida las respuestas de los docentes al sistema de RdC y condenan la reforma al fracaso.

Las políticas de RdC basadas en el desempeño acostumbran a alterar las relaciones de poder existentes en los sistemas de gobernanza escolar (Maxcy, 2009). Estas políticas tienden a otorgar mayores cotas de poder a la administración, los directores o las familias, a la vez que socavan la capacidad de decisión del profesorado. Cuando los docentes perciben que las políticas les desempoderan o restringen su autonomía profesional (por ejemplo, sobre decisiones en el ámbito curricular), estos generalmente adoptan estrategias de resistencia con la que hacer frente a la aplicación de las reformas. Por ejemplo, en Sudáfrica los docentes boicotearon la aplicación de un nuevo sistema de RdC ausentándose de las reuniones, negándose a enviar sus registros o a realizar el trabajo estipulado en el marco del programa de RdC (Mosoge y Pilane, 2014).

Los docentes también pueden adoptar estrategias intermedias de negociación y evasión (Perryman, et al., 2011). Por ejemplo, Palmer y Rangel (2011) encontraron que los docentes de Texas intentaron equilibrar la atención de las necesidades de sus alumnos con las presiones procedentes del sistema de RdC para orientar la enseñanza a los contenidos de las pruebas. Por otro lado, Kostogriz y Doecke (2011) muestran cómo los docentes de Australia fueron capaces de reivindicar su estatus profesional y su autonomía bajo un sistema de RdC, mediante una praxis que, en cierta manera, desafiaba el marco regulatorio de las reformas basadas en estándares. Otras investigaciones han encontrado resultados similares, y han observado cómo los docentes negocian su agencia profesional en el marco de las reformas de RdC y desafían de manera creativa la restricción de su autonomía profesional (véase Hardy, 2014; Osborn, 2006; Robinson, 2012).

En definitiva, las respuestas de los docentes frente a los sistemas de RdC pueden incluir un amplio abanico de reacciones que van desde el conflicto y la resistencia abierta, hasta estrategias de resistencia encubierta y el consentimiento. Dichas respuestas, en lugar de darse por supuesto, deberían conformar un eje fundamental en la investigación empírica sobre los efectos de la RdC.

El contexto social y económico de las reformas de rendición de cuentas

Los contextos institucionales y socio-económicos donde las políticas de RdC son aplicadas condicionan los resultados de dichas políticas. La puesta en práctica de las políticas de RdC no tiene lugar en el vacío sino que se produce en escuelas concretas ubicadas en contextos con importantes condicionantes externos, sociales y materiales. En palabras de Ball et al. (2012, p. 19):

Policies enter different resource environments; schools have particular histories, buildings and infrastructures, staffing profiles, leadership experiences, budgetary situations and teaching and learning challenges.

Así pues, más allá del rol de los actores a título individual, para comprender los efectos de las reformas de RdC deben tenerse en cuenta las interrelaciones entre las dimensiones materiales, sociales y estructurales que configuran las dinámicas de las escuelas, las cuales pueden ejercer presión y actuar como facilitadores o limitadores en el despliegue de las políticas educativas (Ball, et al., 2012).

Muy a menudo, la investigación sobre procesos de mejora y reforma escolar y, en particular, sobre reformas de RdC omite las variables de naturaleza contextual o institucional de las escuelas. Sin embargo, también encontramos excepciones. Por ejemplo, Keddie (2014) realizó un estudio de caso cualitativo en una escuela inglesa ubicada en un contexto socialmente aventajado. Según esta autora, los factores situados, profesionales, materiales y externos mediaron estratégicamente en la puesta en práctica de las políticas de autonomía escolar y de RdC. Entre otras cosas, estos factores posibilitaron a la escuela la adopción de dinámicas cooperativas y le permitieron desafiar la “cultura de la auditoria” así como algunos de los efectos negativos que acostumbran a derivar de ésta. Por otro lado, en base a un estudio conducido en Sudáfrica, Douglas (2005) concluye que los recortes en la plantilla docente, las ratios (alumnos-profesor) excesivamente grandes, y los materiales inadecuados para la enseñanza y el aprendizaje construyeron la puesta en práctica de la reforma del sistema de evaluación del desempeño.

La comparación entre escuelas con diferente rendimiento académico es una manera habitual de intentar capturar la influencia de las variables contextuales y el rol de los factores institucionales en la puesta en práctica de las reformas. En general, la evidencia existente sobre esta temática muestra que las escuelas que, según el sistema de RdC, se encuentran en período de prueba (*on probation*) tienden a adoptar estrategias a corto plazo (más “cosméticas”) para salir de su situación, mientras que las escuelas con alto rendimiento, que no tienen tantas presiones externas, tienden a tener más posibilidades de introducir mejoras en el ámbito de la enseñanza y en el ámbito instructivo. En base a la evidencia recogida en un estudio etnográfico realizado en Chicago, Diamond y Spillane (2004) muestran que el estatus (período de prueba o alto rendimiento) de la escuela influye en la adopción de respuestas diferenciales frente a las presiones procedentes del sistema de RdC. Los hallazgos de esta investigación muestran que las escuelas en período de prueba adoptaron respuestas más superficiales y más reactivas a las sanciones y a las presiones externas, mientras que las escuelas de alto rendimiento alinearon sus respuestas con las aspiraciones reales de los *policy-makers* (por ejemplo, introduciendo cambios en la instrucción e innovaciones). Por supuesto, los factores institucionales y socioeconómicos tienden a estar fuertemente relacionados. En su análisis comparativo de las escuelas de Chicago, Finnigan (2010) observa que los factores socioeconómicos están directamente relacionados con las escuelas

catalogadas como “en período de prueba”, ya que un 92% de las escuelas de Chicago con dicho estatus tienen una mayoría de estudiantes de bajos ingresos.

Las reformas educativas basadas en políticas de RdC de altas consecuencias tienden a favorecer que los docentes más motivados se planteen dejar las escuelas en período de prueba, exacerbando así las desigualdades entre escuelas. En una investigación conducida en once escuelas en período de prueba de Estados Unidos, Mintrop (2003) encontró que los maestros se mostraban escépticos frente a las políticas de RdC y consideraban que el sistema de RdC no reconocía suficientemente su esfuerzo en un contexto educativo con grandes dificultades. Por ello, este autor concluye que el sistema de RdC promovía que los docentes más motivados se plantearan abandonar las escuelas en período de prueba para ir a trabajar a escuelas más aventajadas.

Por su lado, Falabella (2014) examina la influencia del contexto institucional de las escuelas chilenas frente al sistema de RdC. Esta investigadora distingue tres grupos de escuelas (escuelas de alto rendimiento, escuelas de rendimiento medio-bajo, escuelas de bajo rendimiento). En el caso del primer grupo, la autora observa que las políticas de RdC sirven para reforzar el prestigio social de las escuelas y su posición aventajada en el mercado educativo local. Sin embargo, las políticas de RdC no generan incentivos para introducir cambios pedagógicos o de gestión en las escuelas aventajadas. El segundo grupo de escuelas dispone de una demanda muy fiel, normalmente por razones más simbólicas que académicas, de manera que dichas escuelas tampoco tienen incentivos para introducir mejoras educativas en base a los resultados de las pruebas estandarizadas. Finalmente, el último grupo de escuelas, formado por las escuelas de bajo rendimiento, es muy sensible a los resultados del sistema de RdC. Generalmente se trata de escuelas con elevados porcentajes de alumnado vulnerable donde la RdC genera una sensación de estrés debido a la vigilancia y a las sanciones que se asocian a los malos resultados (Falabella, 2014).

Discusión final: Cuatro premisas para futuras investigaciones

Actualmente numerosos países están adoptando políticas de RdC basadas en pruebas de aprendizaje como una forma de mejorar la eficacia de sus sistemas educativos y la productividad

de su profesorado, así como de garantizar unos niveles mínimos de equidad educativa. No obstante, esta forma de RdC se está globalizando a pesar de que todavía hay muchas incertidumbres sobre si ésta funciona o no, en qué sentido y bajo qué circunstancias.

En este artículo, hemos revisado la literatura existente sobre el tema a través del método de la *scoping review*. La adopción de esta metodología nos ha permitido indagar en el amplio espectro de investigación sobre el tema y sistematizar sus resultados más relevantes. Es más, este método nos ha permitido reflexionar sobre cómo los diferentes diseños de los sistemas de RdC, procesos de recontextualización, y contingencias de carácter contextual e institucional pueden conducir a efectos y resultados muy diferenciados. De nuestra revisión se derivan cuatro premisas básicas para la investigación futura sobre la temática. Éstas son:

1) La RdC no es una política educativa específica, más bien abarca opciones de política muy diferentes; diferentes diseños de RdC pueden dar lugar a resultados divergentes, pero también a un amplio abanico de respuestas y procesos de “*enactment*” o recontextualización.

2) A pesar de que la RdC en educación es vista principalmente como una herramienta para mejorar el aprendizaje de los estudiantes, ésta también tiene efectos en otras dimensiones educativas. Algunos de estos efectos son deseados y/o esperados, mientras que otros no lo son.

3) La RdC es generalmente una opción política que genera conflicto y controversia entre la comunidad educativa; el proceso de adopción de sistemas de RdC -así como las características particulares de estos sistemas- condicionan su recepción y puesta en práctica por parte de los docentes, así como sus consecuencias finales.

4) La RdC en educación, como cualquier otra política global, no es resiliente al contexto. Los contextos institucionales y socioeconómicos en los que se promulgan y aplican las políticas de RdC, condicionan la trayectoria y los resultados finales de estas políticas.

Estas cuatro premisas pueden contribuir a sentar las bases de investigaciones sobre RdC más comprensivas y que interpelen a la realidad de diferentes contextos educativos. Al fin y al cabo, estas cuatro premisas nos indican cuáles son los elementos que aún están poco investigados en

la literatura existente sobre la RdC en educación, y apuntan a problemas de causalidad en el estudio de los efectos de este tipo de reformas.

En concreto, la premisa 1 refleja la importancia de examinar las características específicas del diseño de las políticas de RdC como un primer paso necesario para promover evaluaciones más rigurosas de las trayectorias y efectos de dichas políticas. Debido a sus importantes implicaciones, las opciones de diseño no deberían ser vistas simplemente como decisiones técnicas, sino como decisiones con implicaciones profundamente políticas. Como hemos visto, algunas de estas opciones tienen la capacidad de alterar las relaciones de poder fundamentales dentro de los sistemas educativos y generar tipos de respuestas muy diferentes entre los docentes y otros actores educativos clave.

De la premisa 2 se deriva que los efectos no esperados/no deseados de la RdC deben ser también contemplados en los modelos de análisis de estudios futuros. Todavía hay una parte importante de la investigación sobre la RdC en educación que se centra casi exclusivamente en los resultados de aprendizaje como la única variable dependiente. Sin duda, esto es algo que restringe profundamente nuestra comprensión de la naturaleza sistémica de los efectos de la RdC en el campo educativo. Por ello, cabe indagar todavía más sobre los efectos de las medidas de RdC sobre el trabajo y los comportamientos de los docentes, y acerca de qué tipo de sistemas de RdC son más proclives a fomentar actitudes de colaboración (o de competencia) entre los agentes educativos.

La premisa 3 remite a la importancia de abrir la caja de negra que existe entre el diseño de las políticas y sus resultados finales. La implementación de políticas no es un proceso mecánico, sino que un proceso moldeado fuertemente por variables subjetivas (motivación docente, identidades profesionales, percepciones y creencias, múltiples interpretaciones, etc.) que son difíciles de medir en términos únicamente cuantitativos. A pesar de la existencia de un cuerpo emergente de investigación que se centra en este tipo de variables, la investigación que indaga sobre las conexiones entre, por un lado, la respuesta de maestros, directores y otros agentes educativos ante las políticas de RdC y, por otro lado, diferentes tipos de resultados educativos es todavía escasa.

Por último, la premisa 4 refleja la importancia de considerar variables de carácter contextual a la hora de analizar el despliegue y el impacto de las políticas de RdC. A través de nuestra revisión, hemos detectado que la mayoría de investigaciones sobre el tema están sesgadas hacia un enfoque de la eficacia escolar (o *school effectiveness*) que por lo general no presta suficiente atención a cómo el contexto socioeconómico interfiere en cómo las políticas de RdC operan a nivel escolar. La realización de estudios comparativos entre países, estados y/o regiones podría contribuir a poner de relieve la compleja interacción entre entornos institucionales y socioeconómicos, procesos de recontextualización y los efectos de la RdC en múltiples dimensiones educativas, que pueden ir desde el rendimiento académico a la satisfacción y el desarrollo profesional del personal docente.

En general, la forma en que los profesores son concebidos y tratados en el marco de los sistemas de RdC, pero también cómo los docentes interpretan, interactúan y actúan en el marco de estos sistemas es clave para entender los resultados diferenciales - y en ocasiones contradictorios - de la RdC en el campo educativo. La RdC de altas consecuencias tiende a generar más resistencia que otros modelos de RdC, entre otras cosas, debido a los problemas de atribución que plantea cuando trata de definir qué parte del éxito y del fracaso estudiantil es atribuible a los docentes. Autores como Leithwood y Earl (2000) señalan que los sistemas de RdC serían más legítimos si no responsabilizaran a los docentes de forma casi exclusiva de los resultados de aprendizaje de sus alumnos –ya que estos resultados se explican por muchos factores. En lugar de ello, los sistemas de RdC deberían más bien responsabilizar al profesorado de hacer un uso de los recursos disponibles lo más beneficioso posible de cara a avanzar hacia un amplio abanico de objetivos educativos que incluyan objetivos de rendimiento académico, pero también de equidad y de gobernanza democrática de la educación.

Referencias bibliográficas

Alegre, M. A. (2015). Què funciona en educació: la pregunta necessària. Barcelona: Fundació Jaume Bofill & Ivàlua. Recuperado de: http://www.ivalua.cat/documents/1/07_07_2015_15_54_14_Què_funciona_educacio.pdf

- Arksey, H., y O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. doi: <http://dx.doi.org/10.1080/1364557032000119616>
- Andersen, S. C. (2008). The impact of public management reforms on student performance in Danish schools. *Public Administration*, 86, (2), 541-558. doi: 10.1111/j.1467-9299.2008.00717.x
- Anderson, J. A. (2005). *Accountability in education*. Paris: UNESCO-International Institute for Educational Planning.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267. doi: 10.3102/0013189X07306523
- Ball, S. J., Maguire, M., y Braun, A. (2012). *How schools do policy. Policy Enactment in Secondary Schools*. New York: Routledge.
- Boarini, R., y Lüdemann, E. (2009). The role of teacher compensation and selected accountability policies for learning outcomes: An empirical analysis for OECD countries. *OECD Journal: Economic Studies*, 1, 1-20. doi:10.1787/eco_studies-v2009-art10-en
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, 42(2), 231-268. doi: 10.3102/00028312042002231
- Bovens, M. (2007). Analysing and assessing accountability: a conceptual framework. *European Law Journal*, 13(4), 447-468. doi: 10.1111/j.1468-0386.2007.00378.x
- Breakspear, S. (2012). The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance. OECD Education Working Papers, 71. OECD Publishing. doi: <http://dx.doi.org/10.1787/5k9fdfqffr28-en>
- Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, 93(9), 1045-1057. doi:10.1016/j.jpubeco.2009.06.002
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *The Teachers College Record*, 106(6), 1047-1085.

- Day, C. (2002). School reform and transitions in teacher professionalism and identity. *International Journal of Educational Research*, 37(8), 677-692. doi:10.1016/S0883-0355(03)00065-X
- Dee, T. S., y Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446. doi:10.1002/pam.20586
- Diamond, J., y Spillane, J. (2004). High-stakes accountability in urban elementary schools: challenging or reproducing inequality?. *The Teachers College Record*, 106(6), 1145-1176.
- Douglas J. (2005). *Teacher professionalism and education transformation*. Braamfontein: Centre for Education Policy Development (CEPD).
- Escardíbul, O. (2015). Is the introduction of pay-for-performance salary incentives for teachers linked to students' academic performance advisable? What Works in Education Series, 1, Barcelona: Fundació Jaume Bofill & Ivalua. Recuperado de: http://www.ivalua.cat/documents/1/27_10_2015_13_44_00_WhatWorksInEducation_english.pdf
- Falabella, A. (2014). "Do national test scores and quality labels trigger school self-assessment and accountability? A critical analysis in the Chilean context." *British Journal of Sociology of Education*, 37(5), 1-18. doi:10.1080/01425692.2014.976698
- Finnigan, K.S. (2010). Principal Leadership and Teacher Motivation under High-Stakes Accountability Policies, *Leadership and Policy in Schools*, 9(2), 161-189. doi:10.1080/15700760903216174
- Hamilton, L. S., Stecher, B. M., Marsh, J. A., McCombs, J. S., y Robyn, A. (2007). *Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. Rand Corporation.
- Hardy, I. (2014). A logic of appropriation: enacting national testing (NAPLAN) in Australia, *Journal of Education Policy*, 29(1), 1-18. doi: 10.1080/02680939.2013.782425
- Hatch, T. (2013). Beneath the surface of accountability: Answerability, responsibility and capacity-building in recent education reforms in Norway. *Journal of Educational Change*, 14 (2), 113-138

- Hogan, A., Sellar, S. y Lingard, B. (2016). "Corporate Social Responsibility and Neo-Social Accountability in Education: The Case of Pearson plc." In Verger, A., Lubienski, C., & Steiner-Khamsi, G. (eds). *World Yearbook of Education: The Global Education Industry* (pp. 107-124). New York: Routledge.
- Jacob, B. A. y Levitt, S.D. (2003). Rotten Apples: An Investigation Of The Prevalence And Predictors Of Teacher Cheating, *Quarterly Journal of Economics*, 118(3), 843-877. doi: 10.1162/00335530360698441
- Jeffrey, B., y Woods, P. (1996). Feeling deprofessionalised: The social construction of emotions during an OFSTED inspection. *Cambridge Journal of Education*, 26(3), 325-343. doi: 10.1080/0305764960260303
- Jennings, J. L. (2010). School choice or schools' choice? Managing in an era of accountability. *Sociology of Education*, 83(3), 227-247. doi: 10.1177/0038040710375688
- Jones, B. D., y Egley, R. J. (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives*, 12(39), 1-34. doi: <http://dx.doi.org/10.14507/epaa.v12n39.2004>
- Kalimullah, N. A., Ashraf, K. M., y Ashaduzzaman, M. N. (2012). New public management: Emergence and principles. *Bup Journal*, 1(1), 1-22. ISSN: 2219-4851
- Keddie, A. (2014). 'It's like Spiderman... with great power comes great responsibility': school autonomy, school context and the audit culture. *School Leadership & Management*, 34(5), 502-517. doi: 10.1080/13632434.2014.938040
- Keddie, A. (2015). School autonomy, accountability and collaboration: a critical review. *Journal of Educational Administration and History*, 47(1), 1-17. doi: 10.1080/00220620.2015.974146
- Kostogriz, A., y Doecke, B. (2011). Standards-based accountability: Reification, responsibility and the ethical subject. *Teaching Education*, 22(4), 397-412.
- Ladd, H. F. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of Education Review*, 18(1), 1-16.
- Ladd, H. F., y Lauen, D. L. (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management*, 29(3), 426-450.

- Leithwood, K., y Earl, L. (2000). Educational accountability effects: An international perspective. *Peabody Journal of Education*, 75 (4), 1-18.
- Linn, R. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16. doi: 10.3102/0013189X029002004
- Maroy, C., y Voisin, A. (2013). As transformações recentes das políticas de accountability na educação: desafios e incidências das ferramentas de ação pública. *Educação & Sociedade*, 34(124), 881-901.
- Maxcy, B. (2009). New public management and district reform: Managerialism and deflection of local leadership in a Texas school district. *Urban Education*, 44(5), 489-521. doi: 10.1177/0042085908318778
- Mays, N., Roberts, E. y Popay, J. (2001) “Synthesising research evidence”. In N. Fulop, N., Allen, P., Clarke, A., & Black, N. (eds.) *Studying the Organisation and Delivery of Health Services: Research Methods*. (pp. 188-220). London: Routledge.
- McDermott, K.A. (2007). “Expanding the Moral Community” or “Blaming the Victim”? The Politics of State Education Accountability Policy. *American Educational Research Journal*, 44 (1), 77–111. doi: 10.3102/0002831206299010
- Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal*, 30(2), 521–46.
- Mintrop, H. (2003). The limits of sanctions in low-performing schools: A study of Maryland and Kentucky schools on probation. *Education Policy Analysis Archives*, 10(3), 1-30.
- Mosoge, M. J., y Pilane, M. W. (2014). Performance management: the neglected imperative of accountability systems in education. *South African Journal of Education*, 34(1), 1-18. doi: 10.15700/201412120947
- Müller, J. y Hernández, F. (2010). On the geography of accountability Comparative analysis of teachers' experiences across seven European Countries. *Journal of Educational Change*, 11(4), 307-322.

- O’Flaherty, J., y Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education*, 27, 85-95. doi: <http://dx.doi.org/10.1016/j.iheduc.2015.02.002>
- Ohemeng, F., y McCall-Thomas, E. (2013). Performance management and “undesirable” organizational behaviour: Standardized testing in Ontario schools. *Canadian Public Administration*, 56 (3), 456-477. doi: 10.1111/capa.12030
- Osborn, M. (2006). Changing the context of teachers’ work and professional development: A European perspective. *International journal of educational research*, 45(4), 242-253. doi:10.1016/j.ijer.2007.02.008
- Palmer, D., y Rangel, V. S. (2011). High stakes accountability and policy implementation: Teacher decision making in bilingual classrooms in Texas. *Educational Policy*, 25(4), 614-647. doi: 10.1177/0895904810374848
- Perryman, J., Ball, J.S., Maguire, M., y Braun, A. (2011). Life in the Pressure Cooker – School League Tables and English and Mathematics Teachers’ Responses to Accountability in a Results-Driven Era, *British Journal of Educational Studies*, 59(2), 179-195. doi: 10.1080/00071005.2011.578568
- Powers, J. M. (2003). An analysis of performance-based accountability: Factors shaping school performance in two urban school districts. *Educational Policy*, 17(5), 558-585. doi: 10.1177/0895904803256789
- Ranson, S. (2003). Public accountability in the age of neo-liberal governance, *Journal of Education Policy*, 18,(5),459-480. doi: 10.1080/0268093032000124848
- Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92(5), 1394-1415. doi: 10.1016/j.jpubeco.2007.05.003
- Reback, R., Rockoff, J., y Schwartz, H. L. (2014). Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under No Child Left Behind. *American Economic Journal: Economic Policy*, 6(3), 207-241. doi: 10.3386/w16745

- Robinson, S. (2012). Constructing teacher agency in response to the constraints of education policy: adoption and adaptation. *Curriculum Journal*, 23(2), 231-245. doi: 10.1080/09585176.2012.678702
- Roderick, M., Jacob, B. A., y Bryk, A. S. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation and Policy Analysis*, 24, (4), 333-357.
- Rosenblatt, Z., y Shimoni, O. (2001). Teacher accountability: An experimental field study. *Journal of Personnel Evaluation in Education*, 15 (4), 309-328.
- Ryan, J. E. (2004). Perverse Incentives of the No Child Left behind Act, *The New York University Law Review*, 79, 932-989.
- Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, 11 (1), 45-61.
- Tobin, M., Lietz, P., Nugroho, D., Vivekanandan, R., y Nyamkhuu, T. (2015). *Using large-scale assessments of students' learning to inform education policy: Insights from the Asia-Pacific region*. Melbourne: Australian Council for Educational Research (ACER).
- UNESCO. (2015). Framework for Action Education 2030: Towards inclusive and equitable quality education and lifelong learning for all. Recuperado de: <http://www.uis.unesco.org/Education/Documents/wef-framework-for-action.pdf>
- Volante, L. (2004). Teaching to the Test: What Every Educator and Policy-Maker Should Know. *Canadian Journal of Educational Administration and Policy*, 35. Recuperado de: <http://www.umanitoba.ca/publications/cjeap/articles/volante.html>
- West, A., Mattei, P., y Roberts, J. (2011). Accountability and Sanctions in English Schools, *British Journal of Educational Studies*, 59 (1), 41-62. ISSN 0007-1005
- World Bank. (2015). What Matters Most for School Autonomy and Accountability: A Framework Paper. SABER Working Paper Series, 9. Recuperado de: https://olc.worldbank.org/sites/default/files/Whatmatters_SAA_framework_paper_Demas_2014_4.pdf

Apéndice

Tabla 1.

Sintaxis

```
TITLE-ABS-KEY ("education"OR "education reform"OR "School")AND TITLE-ABS-KEY("Effective school research"OR "Efficiency"OR "Effectiveness"OR “Effects” OR “Outcomes”)AND TITLE-ABS-KEY("New Public Management"OR "NPM"OR "School autonomy"OR "Institutional autonomy"OR "Accountability"OR “Merit-based pay” OR “School-based management”OR “Public management reform”) AND NOT ("Higher education") AND ( LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) OR LIMIT-TO(SUBJAREA,"ARTS" ) ) AND ( LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) ) )
```

```
TITLE-ABS-KEY ("education"OR "education reform"OR "School")AND TITLE-ABS-KEY(“Enactment”OR “Implementation”)AND TITLE-ABS-KEY("New Public Management"OR "NPM"OR "School autonomy"OR "Institutional autonomy"OR "Accountability"OR “Merit-based pay” OR “School-based management”OR “Public management reform”) AND NOT ("Higher education") AND ( LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) OR LIMIT-TO(SUBJAREA,"ARTS" ) ) AND ( LIMIT-TO(SUBJAREA,"SOCI" ) OR LIMIT-TO(SUBJAREA,"BUSI" ) OR LIMIT-TO(SUBJAREA,"ECON" ) ) )
```

Fuente: autores

Agradecimientos

Este trabajo ha sido realizado en el marco del programa de doctorado en Sociología de la Universidad Autónoma de Barcelona y ha contado con el apoyo de la beca FPU del Ministerio de Educación, Cultura y Deporte español (ref. FPU13/01090). Asimismo, los autores también agradecen el financiamiento entregado por el European Research Council bajo el programa marco Horizon 2020 de investigación e innovación de la Unión Europea, número de subvención 680172 - REFORMED: <http://reformedproject.eu>.

Publication D: Verger, A., Ferrer-Esteban, G., & Parcerisa, L. (2020, forthcoming). In and out of the 'pressure cooker': Schools' varying responses to accountability and datafication. In: S. Grek, C. Maroy & A. Verger (eds.) *World Yearbook in Education 2021: Accountability and datafication in education*. New York: Routledge.

CHAPTER 4

In and out of the ‘pressure cooker’: Schools’ varying responses to accountability and datafication

Introduction⁷

In educational research, performance-based accountability (PBA) systems have often been likened to ‘pressure cookers’ (cf. Agrey, 2004; Perryman et al., 2011; Tan, 2018). PBA puts high levels of pressure on schools by holding them liable for their performance, sanctioning underachievement and rewarding success. With high-stakes accountability systems in particular, underperforming schools experience higher levels of pressure, since continuous low performance has significant implications, from having restrictions placed on schools’ pedagogic, organizational and economic autonomy to being forced into closure (Diamond, 2012; Kim and Sunderman, 2005; Mintrop, 2004). By exerting these and other types of pressures, PBA is expected to make schools more responsive to the achievement of centrally defined learning goals and more inclined to use learning metrics in their daily practices and decisions. Overall, PBA policies aim to schools more consciously aligning their instruction with the mandated curriculum, and more intensively using achievement data to identify learning gaps and define educational and organizational improvement programs. Through the promotion of these changes, PBA systems not only aim to improve learning achievement in aggregated terms but also to ensure that schools (especially underperforming schools) pay sufficient attention to their most disadvantaged students.

Although the expectations with PBA are high, existing research shows that the impact of PBA on school organizations is rather uneven. Accountability instruments can generate a broad range of outcomes and responses, from altering the goals and organizational identity of schools to more short-term strategies and cosmetic changes, such as intensifying teaching to the test (i.e., focusing instruction on predictable test content and/or test formats) or narrowing the curriculum (i.e., dedicating more time and resources to tested subjects) (Au, 2007; Mittleman

⁷ This work was supported by the European Research Council under the European Union’s ‘Horizon 2020 Framework Programme for Research and Innovation’ [GA-680172 – REFORMED].

and Jennings, 2018). In the latter case, rather than aligning school practices with accountability expectations, such responses are more likely to decouple from them. Varying school responses to PBA are identified in both high-stakes and low-stakes accountability systems, as well as in countries where these systems have been installed for shorter or longer periods of time (see Diamond and Spillane, 2004; Landri, 2018; Maroy and Pons, 2019; Candido, 2020). Even in the US context, where decades of high-stakes accountability have turned schools into testing- and data-intensive organizations (see Mittleman and Jennings, 2018), schools' pedagogic and organizational responses are far from homogeneous.

Existing research tends to attribute the varying responses to accountability pressures to variables of a different nature, ranging from school leadership styles to the broader socioeconomic contexts in which schools operate. However, so far, research has overlooked the role of subjective variables (such as school actors' perceived and experienced pressures) in the mediation and enactment of PBA. To address this gap, this chapter aims to analyze the production of different patterns of responses to PBA within schools from a policy enactment perspective. On the basis of a mixed-methods study conducted in Chile, we analyze how school actors' interpretations of and dispositions toward PBA, on the one hand, and their experienced levels of pressure, on the other, influence how they respond to the accountability regulatory system. As we will show, the responses to PBA that have been identified go beyond conventional alignment-decoupling dichotomy and include a more varying range of options. Our perspective is premised on the assumption that the way school actors respond to policy prerogatives is contingent on the way these actors make sense of PBA pressures and expectations within their broader social and institutional frameworks. In other words, the responses to PBA that we identify are the result of analyzing how school actors *see* and *live* accountability regulations in their reference contexts.

To build our main arguments, the chapter is structured as follows: in the first section, on the context of the research, we introduce Chile's long trajectory of experimentation with learning metrics and a broad range of related accountability measures. In the second section, we present our theoretical framework, where we highlight the importance of focusing not only on policy interpretation but also on perceived regulatory pressure to understand how policies are enacted. After presenting the methodology of our study in the third section, in the fourth one, we offer

the main findings of the research in the form of a new categorization of school responses to PBA regulations. Finally, the conclusions highlight the key mediating role of subjective variables in the configuration of different patterns of school responses to PBA, and we reflect on the research and policy implications of our study.

Context of the study: Governing schools through performativity and markets in Chile

Chile is a country where PBA has a long tradition in the governance of education. Chile was an early adopter of national, large-scale assessments in education, the first of which was implemented at the start of the 1980s, in the context of an ambitious and drastic market reform promoted by the civic-military dictatorship (1973–1990). At that time, the main intention of the national assessment, known as the System for the Measurement of Quality in Education (SIMCE), was to inform school choice (Bellei and Vanni, 2015). Nevertheless, it was not until restoration of democracy in the 1990s (specifically, in 1995) that SIMCE results started being publicly disseminated on a regular basis (first in the media and later on the Ministry of Education website). Since then, more and more functions and uses have been found for this standardized test. To start with, in 1996, the SIMCE became a fundamental component of a merit-based pay system for teachers, in which collective salary bonuses were attached to schools' performance (Mizala and Schneider, 2014).

During the early 2000s, the government implemented various compensatory programs aimed at low-performing schools, to promote data use and school improvement processes (Falabella, 2020). This policy approach crystallized in the enactment of the Preferential School Subsidy Law in 2008. Under this scheme, the state gave an additional subsidy to schools for each 'vulnerable' student enrolled. As a condition of accessing this subsidy, schools accepted additional accountability measures. Schools were classified according to their SIMCE performance, and in the case of continuous poor performance, the state could impose sanctions which included the possibility of school closure (Valenzuela and Montecinos, 2017).

In 2011, a new Education Quality Assurance System (Law n° 20529) was created, whose provisions allow the Chilean state to adopt new data-intensive policy instruments and tools to

inspect, evaluate and sanction all types of public and publicly subsidized private schools – not only those receiving the preferential school subsidy (Parcerisa and Falabella, 2017). Since then, schools have been classified in four performance categories (high, medium, medium-low and insufficient) according to SIMCE results, data on students' learning progress, and a set of personal and social development indicators.⁸ Poorly performing schools are meant to receive pedagogical support and external evaluative visits from the Ministry of Education for a period of four years, and if their performance remains insufficient, schools can be closed. In parallel, the Education Quality Agency (EQA) has put a great deal of effort into making performance data intelligible and actionable for the elaboration of school improvement plans through various initiatives, online tools and training seminars.

In short, Chile is a country where both performativity and datafication, in interaction with market rules, have a great potential to alter school practices and to discipline teachers' behavior. Chilean education is a distinctive scenario where to study the combined effects of market forces and different forms of administrative accountability pressure on educational organizations and practices.

Understanding the variation in school responses through enactment theory

Interpretation as a key moment in policy enactment

Enactment and sense-making theories are well suited to exploration of how school actors 'construct the demands of, and appropriate responses to, accountability systems differently' (Jennings, 2010, p.229). Such theories broadly state that the way educational actors interpret and make sense of new policy mandates is key to explaining how such mandates translate into everyday practices (Ball et al., 2012; Spillane et al., 2002). These theories do not portray teachers and principals as simple policy takers but as policy shapers who actively adapt, modify and reframe new policy prerogatives to suit their preferences and the needs and constraints of their particular school contexts. Following a policy enactment approach, school responses to policy instruments such as PBA will 'depend on how the aims and purposes ascribed to them, and the meanings and representations they carry, are perceived [and] understood [...] by key actors'

⁸ <https://www.agenciaeducacion.cl/se-obtiene-la-categoria-desempeno/>

(Skedsmo, 2011, p.7). This does not mean that policy interpretation is mainly guided by instrumental rationality and causal beliefs. Principled beliefs, personal biographies, previous experiences or emotional scripts co-constitute the interpretive frames through which educators approach educational policy, and respond to it. Furthermore, policy interpretation, beyond an individual act, results from the interaction between school actors, and within a broader network that includes parents, the school owner, inspectors and external consultants, among others (Spillane et al., 2002).

Policy interpretation is key to understanding why some schools align themselves with new policy reforms, but others avoid implementing them. According to Malen (2006), school actors align with new reforms when they perceive that these reforms easily couple with their previous way of working and/or their particular or collective interests, whereas school actors with a more conflicting approach to reform implementation are those that disagree with the main goals and/or instruments of educational reform. School actors might dislike an educational reform due to concerns about its usefulness, validity or fairness, or out of concern that it goes against their interests (or the interests of students) or contradicts their professional values and educational beliefs. When negative interpretations predominate, schools tend to address the external pressure to comply with new regulations through dilution strategies and obstructive bureaucratic games, such as neglect, overt resistance or subtle adaptation (Malen, 2006). Some of these dynamics have been observed in several Francophone countries, where teachers justify ritualistic (but not substantive) adoption of accountability instruments because they consider that these instruments clash with their own notions of good instruction and student assessment (see Maroy et al. in this volume).

The role of actors' subjective perceptions of accountability pressures

PBA systems assume that the more pressure is exerted on schools, the more measures will be taken to enhance the educational quality of school provision. Pressure is likely to be exerted based on schools' performance levels; thus, schools on probation will be subjected to stronger and more coercive forms of accountability pressure. High levels of pressure may result in different responses, typically defined as alignment or decoupling, depending on whether the accountability expectations are met or not - with the latter usually being associated with tactical

or symbolic responses in order to cope with regulatory pressures, at least in the short term (cf. Boxenbaum and Jonsson, 2016).

Research conducted in high-stakes accountability settings often concludes that schools exposed to coercive sanctions – which, not coincidentally, are those that tend to serve the most disadvantaged student populations or concentrate on students who are more challenging to teach (see Pallas in this volume) – are those that more frequently adopt instructional tactics through which to inflate test results, without necessarily adopting ‘deep’ changes in pedagogy (Mittleman and Jennings, 2018, p.481). Thus, schools facing higher levels of administrative pressure tend to intensify test preparation practices and the number of teaching hours dedicated to subjects evaluated externally; to focus further on so-called ‘bubble students’ (i.e., students who are closer to the proficiency cut-off score); or to track students according to their performance level in order to customize their training (Au, 2007; Mittleman and Jennings, 2018; Watanabe, 2008).

Accountability research typically assumes that pressure is equally high (once school performance is held constant), while contextual factors foster or hinder full implementation of school improvement policies. Specifically, instrumental responses may depend on factors such as school composition or performance levels (which, in turn, are related to composition). Schools’ socioeconomic and institutional characteristics define the limits of what is possible and desirable in terms of school improvement and performativity, and how much pressure PBA exerts on them. For example, in deprived schools (where aggregate performance levels are likely to be lower and pressure is likely to be higher), school actors will be more prone to carrying out superficial strategies to increase scores in a short amount of time. Research also shows that privately managed schools tend to be more reactive to administrative accountability pressure than conventional public schools and articulate more instrumental responses (Berends, 2015; Zancajo, 2020).

Although we acknowledge that all these dynamics do indeed occur in PBA systems, we also assume that the reality of schools is highly complex, and it cannot be taken for granted that objective school characteristics alone determine the performative pressure that school staff experience. In our view, the intensity of administrative accountability pressures, which are objective in nature, is not constant as might be subjectively mediated. What makes schools

reactive to the regulatory framework is not only the level of pressure that regulations and authorities exert (whether schools are put into the ‘pressure cooker’ or not) but also the pressure that school actors perceive, live and experience. School actors’ perception involves factors which, together with the characteristics of the school, play an important role in explaining divergent school responses.

But on which factors does the intensity of the perceived pressure depend? High-performing schools, and/or schools whose educational provision has a clear focus on academic excellence, will readily align with external PBA demands, so they do not necessarily perceive the PBA system as a source of pressure (see Keddie, 2013). Nonetheless, there might be teaching staff in well-performing schools who feel higher-than-expected performative pressure because their ‘significant others’ (i.e., their more direct competitors) are doing better than they are; or the school owner, the principal, families and/or teachers themselves might think that the academic results of the school have room for improvement. At the same time, schools with poor levels of performance may have staff who do not experience high levels of performative pressure due to the fact that (for moral, professional or pedagogic reasons) they do not put academic achievement at the center of their work.

In short, school responses to regulatory environments are the result of complex policy enactment processes in which variables of a contextual and subjective nature interact. From a sense-making perspective, it follows that two subjective variables are particularly significant in terms of understanding different patterns of school responses to accountability pressure. The first is how school actors interpret and position themselves in the PBA policy debate, and the second one is the perceived level of performance pressure that school actors experience within PBA frameworks.

Methodology

Data

The data used in this research have been drawn from the REFORMED project database, which applies to a sample of countries, including Chile.⁹ We have also used secondary administrative data provided by the Chilean educational authorities.

The Chilean database includes data collected from questionnaires administered to teachers (n = 1130) and school leaders (n = 200), distributed among 79 schools that were selected through a stratified sample strategy (Ferrer-Esteban, 2020). These questionnaires provide rich information about school contexts, the professional profiles of teachers and principals, school organization, teaching practices, use of standardized test data, perceptions of the PBA system and other teachers' beliefs (Levatino, 2020).

Method

The research follows a sequential mixed-methods design approach (cf. Teddlie and Tashakkori, 2006), which integrates two different empirical stages. The *first stage is eminently quantitative*. In this stage, we constructed school categories based on variables related to both attitudes and beliefs about PBA, and perceived performance pressure (as introduced in our theoretical framework). These school categories, in combination with school composition variables (socio-economic composition and performance), were then used to construct the school sample for the qualitative analysis.

During the *qualitative research stage*, we conducted interviews with teachers, principals and other school leaders. This phase was essential in order to identify (using the interpretative framework for the school categories) all those school-level practices that actors may adopt to deal with the accountability system. The manner in which these school practices are enacted (and the intensity) within each category allowed us to characterize and define varying school responses to performance-based accountability.

Quantitative research stage: Constructing school categories

⁹ See www.reformedproject.eu.

This stage was carried out with the entire Chilean sample of schools ($n = 79$). The survey responses of all school actors involved (school leaders and teachers) were considered.

The first step was to identify broad categories of schools on the basis of attitudes and beliefs about PBA and perceived performance pressure. To construct composite indexes of PBA beliefs and performance pressure, we first identified the most significant variables in our survey database in order to capture each of these constructs through factor analysis. As a result of this analysis, the *PBA attitudes/beliefs* index is based on three variables related to both the perceived fairness and validity of the PBA system, as follows: a) whether it is fair to measure school quality through the results of standardized tests; b) whether it is fair to disseminate test results in the media and/or internet; and c) whether test scores reflect the efforts and ability of teachers. The index of *perceived performative pressure* includes variables relating to pressure from different account-holders: the Ministry of Education, the EQA and the municipality (public schools) or school board (private schools).

From the intersection of the two indices, we then defined quadrants, which were used to frame the surveyed schools. Table 1 shows the frequencies of teachers and schools according to school categories, each of which covers between 22% and 30% of the sampled schools.

Table 1. Schools and teachers by performative pressure and culture

<i>School categories</i>	Teachers		Schools	
	<i>Freq.</i>	<i>Perc.</i>	<i>Freq.</i>	<i>Perc.</i>
High pressure and con-PBA	322	28.5	24	30.4
High pressure and pro-PBA	262	23.2	18	22.8
Low-pressure and con-PBA	249	22.0	18	22.8
Low pressure and pro-PBA	297	26.3	19	24.1
<i>Total</i>	1130	100	79	100

Source: Reformed database. Chile

Qualitative research stage: Characterizing school responses

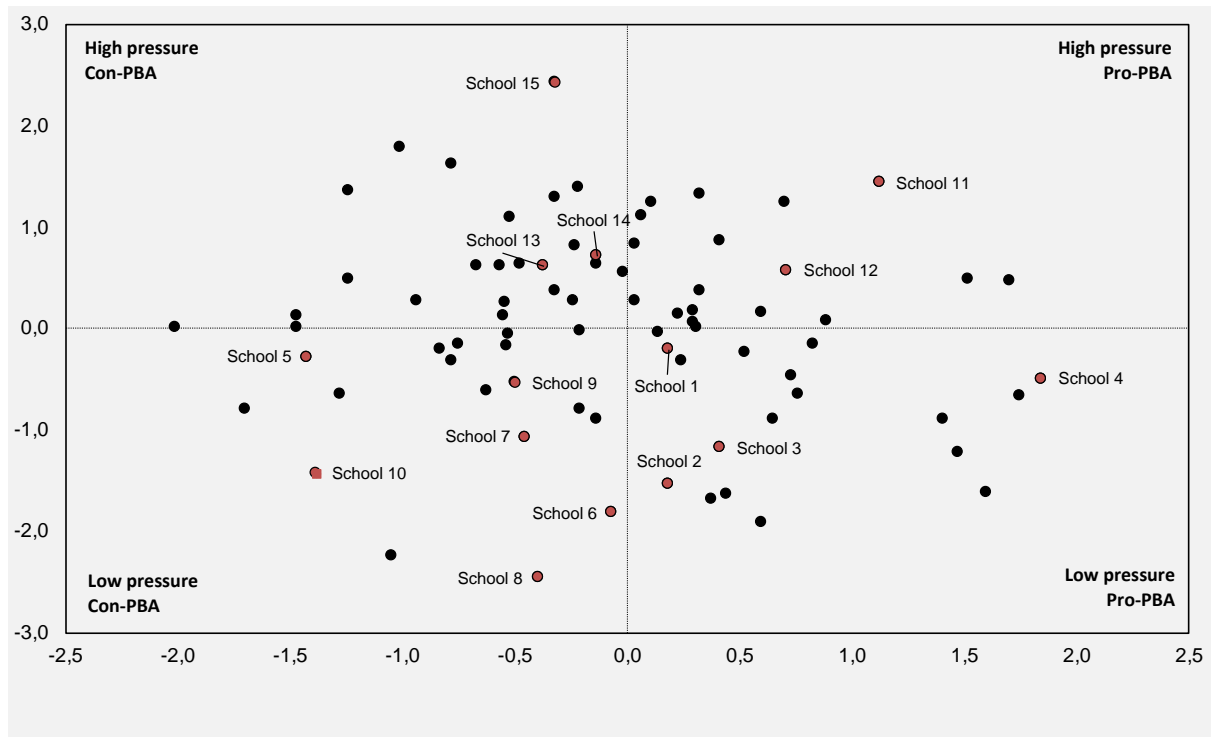
Qualitative fieldwork was conducted using a smaller sample of schools via semi-structured interviews with teachers and school leaders, covering similar topics to those of the survey. The schools for this stage were first selected taking into consideration their social composition (school vulnerability index, MINEDUC, 2019) and level of performance (Agencia de Calidad de la Educación, 2018). In Table 2, we can see how the selected schools for this stage related to our subjective categories (perceived pressure and beliefs about PBA), along with the indicators of social composition and performance. From Figure 1, it can be seen that the same schools are spatially distributed across the quadrants.

Table 2. Schools of the qualitative stage

School	School SES	Performance category	Pressure / PBA approach
1	Low SES	Low-Medium	Low pressure, pro-PBA
2	High SES	High	Low pressure, pro-PBA
3	High SES	Medium	Low pressure, pro-PBA
4	Med-high SES	High	Low pressure, pro-PBA
5	Low SES	Insufficient	Low pressure, con-PBA
6	Med-high SES	Low-Medium	Low pressure, con-PBA
7	Med-high SES	Low-Medium	Low pressure, con-PBA
8	High SES	High	Low pressure, con-PBA
9	Med-high SES	Medium	Low pressure, con-PBA
10	Low-med SES	Low-Medium	Low pressure, con-PBA
11	Low-med SES	Low-Medium	High pressure, pro-PBA
12	Low SES	Medium	High pressure, pro-PBA
13	Low-med SES	Low-Medium	High pressure, con-PBA
14	Low-med SES	Medium	High pressure, con-PBA

Source: Reformed database. Chile

Figure 1. Distribution of schools according to school responses



Source: Reformed database. Chile

Our typology of school responses was mainly defined and constructed through interview data, although administrative data and data from the survey were also used. The tables in the Appendix show more detailed information about the schools' main characteristics (namely, school ownership, socioeconomic composition and performance) by quadrant but also in terms of PBA-related school practices (with a focus on data use and teaching to the test). The interviews enabled us to reconstruct and capture the narratives, rationales and logics of action that predominate in the different groups of schools, and to build patterns of responses to PBA accordingly.

All the interviews with principals and teachers were transcribed and analyzed with QDA software. To analyze the qualitative material, we combined the use of emerging and pre-established codes for key factors, the most relevant of which are as follows: school context and culture, leadership style, actors' opinions about PBA, knowledge about the quality assurance system, the importance given to test results, data use, test preparation strategies, classroom management practices, external support and teacher autonomy.

Findings: School responses to accountability regulations

In this section, we present the most defining characteristics of our school categories, with a focus on depicting the responses to PBA that were found to predominate in each category. The resulting typology of school responses includes induced alignment, accommodation, fabrication, dilution and *de facto* opting out. Conceptually, this typology draws on previous categorizations of educational organizations' responses to regulatory pressures (see Coburn, 2004; Malen, 2006; Landri, 2018).

High pressure and pro-PBA: Induced alignment

In schools with a high level of perceived pressure and a more positive approach to PBA, we found that teachers and principals proactively embrace the mandate of PBA and put data-intensive instruments at the center of the governance system of the school. The academic requirements for these schools tend to be high. In terms of the student population, schools in this category are predominantly 'middle class': almost three quarters of their students are distributed between the second and third SES quarters (see Table a2). Both public and private schools are represented in this category, but private subsidized schools predominate. Specifically, 66.7% of the schools in this category are private subsidized schools.

These schools devote significant effort to incorporating the accountability mandate into not only instructional strategies but also broader pedagogic and management approaches. They do so not necessarily because they need to improve their learning achievement data (in fact, only 5.6% of the schools in this category obtain low results) but because they are externally pressured to

improve their performance and/or reputation in the reference school market. This is why we describe the predominant response in this category as *induced alignment*.

Performative pressure is inflicted by the school owner to a great extent. As the principal of one private subsidized school pointed out, the school owner pressures them ‘to achieve better results’ on the SIMCE test because one of the central missions of the corporation is that its schools become ‘top performing institutions’. Nonetheless, rather than being seen as stressful or frustrating, these pressures are regarded as ‘necessary tensions’ that trigger school change and activate internal improvement dynamics (principal, school 12).

Schools in this category tend to attach a lot of importance to SIMCE results and use them to identify learning needs and school improvement challenges. Almost 50% of the teachers in this category were found to be very positive about how much guidance the SIMCE gives them to help improve the quality of their teaching, while only 21.5% reported that test results are not used or are seldom used (see Table a5). Leadership teams and teachers use learning metrics as benchmarks and planning tools, and articulate and elaborate discourse on the importance of data-management practices.

Look, actually, I think that they [SIMCE results] have affected my work from the perspective that they make an important point; you have to think about them. But you know that what I [get] from SIMCE is that when I see the distribution of results, I say: ‘How many students are not learning?’ and in reality, what drives me there is that this percentage or this number decreases. In my opinion, they are a new impulse to say: ‘How do we make [that number] go down? How do we do it?’ These [results] illuminate my tasks during the year, more than anything else. (pedagogical coordinator, school 12)

In the schools where induced alignment predominates, test preparation practices are intensive, systematic and routinized. Teaching to the test is not only a remedial activity but an educational practice that permeates the educational dynamic of the entire school. Of the teachers in this category, 61.2% reported that they not only prepare their students for the test beforehand but throughout the whole year (see Table a4). Teaching to the test is a common practice across different grades (beyond the SIMCE-assessed grades) and is strategically supported and reinforced by external testing services.

We are working on the SIMCE courses with specific support. Ok? We work in the mode of the SIMCE questions, which means, let’s say,... in certain ways..., preparing them (the students) to

answer in a better way, ok?, in the courses that are evaluated by the SIMCE each year, which are always the 4th grade, and it varies in the 6th and 8th grades. Ok? [But it] is not only the teacher of the 4th grade; all teachers support, let's say, these reinforcements, in the collaborative work that... as I told you, characterizes us as a school. (management team, school 11)

To sum up, schools' alignment with PBA in this category results from the coupling between an academic-oriented school ethos and an externally induced climate of pressure to improve school results. These are schools that regard learning achievement (and high academic standards) as being one of the main focuses of their educational provision.

Low pressure and pro-PBA: Easily accommodating accountability requests

Schools with a positive attitude towards PBA and which do not feel a high level of performative pressure are those that accommodate PBA regulations more readily. The schools in this category are predominantly private (68.4% are privately owned), and they enroll socially advantaged students (42.1% are in the first top SES quarter) (see Tables a1 and a2). Teachers have high expectations of their students and prepare them to access the best universities in the country. In terms of performance, 84.2% of these schools are distributed among the highest categories of achievement (see Table a3). Accountability regulations are convenient for these schools; rather than generating pressure, the regulations reinforce their educational and teaching approach. Thus, these schools easily accommodate the PBA regulations.

Order and discipline, and a culture of effort and academic rigor are the main hallmarks of accommodating schools. These values are instilled into teachers by the school management team and, in turn, are instilled into students by teachers. These schools have a culture of continuous improvement and aim to boost their scores year after year. Nonetheless, their improvement plans can also be holistic, integrating goals and dimensions which go beyond learning achievement. Decisions about teaching strategies, materials and methods are centralized at school level. Some teachers in these schools miss the level of autonomy they had in previous placements and consider their job to be demanding and competitive. However, they also appreciate having clearly defined targets and goals, frequent meetings with school staff and regular feedback from the management team.

Test preparation is a common practice (with 55% of teachers in this category stating that it takes place throughout the year), but it is not necessarily seen as a strategy mainly focused on boosting test scores (see Table a4). Teachers in these schools realize that ‘reinforcement activities’ are more than just ‘SIMCE simulations’ (not fully focused on the SIMCE) and include, for instance, elaboration and delivery of lesson plans inspired by the most challenging SIMCE questions (teacher, school 3). They also do not feel the need to prepare for the SIMCE explicitly because doing well in the SIMCE ‘should be like a natural process’ if they do their job properly throughout the year (vice-principal, school 2).

Data use is highly routinized in these schools: 52% of teachers stated that SIMCE results guide all or most of their teaching, while 28% stated that SIMCE results are useful to a certain extent (see Table a5). Managers and teachers make numerous decisions at school level on the basis of performance data, not only to find room for improvement contingently but to address more structural problems that might have been generated in previous years or in different subjects to those assessed by the SIMCE.

Accommodating schools are very positive about SIMCE data. They trust in the SIMCE’s validity and believe that it provides a trustworthy representation of actual school performance. Staff in these schools feel that if anything, the SIMCE should provide even more information and details, and they argue that they would like more time to analyze and discuss the data, and take further advantage of this information. Overall, accommodating schools have a very elaborated discourse on data use, and they can give numerous examples of how they use data in their everyday work.

We use all the actions recommended by the system. The SIMCE tells us how to focus the question; that is, when we ask a question at an appropriate level, when we ask too [easy] questions... there are teachers who use the tests as feedback; deliver the tests to the students [...]; then the teacher works [out] the questions. This works as a pedagogical space rather than as a learning or evaluation space... (pedagogic coordinator, school 1)

Low pressure and con-PBA: Between dilution and opting out

In the category of schools that perceive a low level of performative pressure and are openly critical about PBA, we found two distinct types of response: dilution and *de facto* opting out. These responses emerge in two drastically different school contexts, which is why it is difficult

to characterize this category in terms of school performance, socioeconomic composition and ownership. However, schools in this category share a low level of test-data use and a very low frequency of teaching to the test (with more than 50% of teachers in this category stating that they never conduct this practice) (see Tables a4, a5, a6 and a8).

a) Dilution

Dilution emerges in schools that feel a low level of performance pressure and do not put academic performance and metrics at the center of their educational approach. The educational provision of these schools tends to focus on critical thinking, students' personal development and/or socio-emotional skills. Some of these schools define themselves as 'revolutionary', as in the case of school 6, whose principal is proud of having adopted a 'circular and integrative neuropsychological educational approach'. This principal made it clear that he is not averse to numerical data but believes achievement data to be too narrow and limited to inform the most important school decisions. He has created his own data governance and assessment tools, which he considers to be more comprehensive and aligned with their pedagogical approach and beliefs.

Managers and teachers in these schools openly avoid using SIMCE data for competition purposes because they believe that quality teaching cannot be developed under 'stress' and when students are put under 'high pressure' (teacher, school 7). Test simulations are not a frequent practice either. The owner and/or management teams of these schools do not ask teachers to intensify teaching to the test before implementation of the SIMCE, but some teachers nonetheless take the initiative to implement test simulations before the SIMCE as a way to familiarize their students with the rules and procedures associated with standardized test conditions. These teachers argue that they provide test preparation out of consideration for students' well-being on the SIMCE day but not to inflate the results. Accordingly, test simulations are low-intensity and mainly happen in relation to SIMCE grades and subjects.

I: [A]lthough it is not something that... they demand from me, we do a couple of essays as a matter of protocol, [so] they know who Ruth [is], that she has to sit in a [certain] way, that you have to fill out an answer sheet, because it is not usual for us to use [these tests] here. (teacher, school 8)

Overall, the schools where PBA pressures are diluted are well placed in the local education market, and their educational provision and their expectations are not mainly concerned with academic achievement. The combination of these factors counteracts performative pressures from the regulatory environment and allows them to develop educational, governance and data-management practices that they think better serve their educational vision.

b) De facto opting out

De facto opting out also emerges in schools that have low levels of perceived administrative pressure and are also critical of the accountability system. However, the logic of action and socio-material characteristics of the schools articulating this response are entirely different from what we saw in relation to policy dilution. Opting-out schools perform worse and enroll much higher rates of disadvantaged students. Rather than emphasizing alternative pedagogy, their educational approach emphasizes the importance of 'living together' and 'security'. These schools are located in disadvantaged neighborhoods where violence is a common concern among both parents and school staff. Families in these predominantly public schools do not have high academic expectations, and what they most value is the security that the school can give to their children.

P: I believe that [the increase in student enrollment in our school] has to do with [our reputation] that... parents can rely on... [F]or example, they bring their children to the school with two eyes, a nose, and all his/her extremities, and at the end of the day we will deliver the same child back to them. I think that this creates confidence. (principal, school 5)

Principals and teachers feel powerless in the face of the improvement requirements coming from the educational authorities. A central characteristic of this type of school is a low level of academic ambition, framed by the perception that given the socioeconomic disadvantages of their students, there is no room to reverse poor SIMCE performance. Contrary to PBA expectations, in these institutional contexts, the threat of school closure does not translate into higher levels of performative pressure.

These schools rarely enact test preparation activities. Management teams do not advocate teaching to the test, and if test simulations are conducted, it is because teachers have decided autonomously to do so. For instance, one teacher told us that she voluntarily took the initiative

to conduct a few test simulations before the SIMCE but mainly for the purpose of familiarizing students with how answers are marked in standardized tests.

Interviewer: Is the decision to not do SIMCE simulations yours or...?

Teacher: Yes, it's mine.

Interviewer: ... or is it school policy?

Teacher: No, it was mine; it was mine with my PIE [School Integration Program, in English] partner [...] I think we did a test simulation, but it was [for the purpose of] marking the answers. (teacher 2, school 5)

Another feature of these schools is a certain disaffection with learning metrics. School staff stated that data from standardized tests are decontextualized and misaligned with their educational focus, and that their professional judgement capacity as educators and related tools (e.g., 'class evaluations') (teacher 2, school 5) are more reliable than test data.

To conclude, *de facto* opting out should not be confused with overt or even covert forms of resistance. Far from being a type of conscious resistance or a policy dilution strategy, *de facto* opting out is a common non-reaction among schools operating in highly marginalized contexts, which feel powerless and resigned in front of unattainable improvement expectations. This response is thus characterized by indifference and passivity in the face of the external pressures generated by the quality assurance system.

High pressure and con-PBA: Fabricating better scores

Schools that experience high levels of performance pressure and are critical of PBA react to accountability regulations superficially and fragmentedly. Within this category, there are three main contextual features that stand out: a clear predominance of public schools (62.5%); low levels of educational performance (with 62.5% of schools in the low and mid-low achievement categories); and an overwhelming presence of disadvantaged students (with more than 90% of the students in these schools coming from the bottom and second quartiles) (see Tables a1, a2 and a3). Despite the performance pressures faced by these schools and experience from different sources (educational authorities, school owner, etc.), they do not transform their core practices (namely their educational provision, their evaluation systems and/or their governance structures) as a response to such pressures. Instead, they adopt various practices and strategies to overcome

accountability pressures but in a way that does not substantively alter their educational values and approach.

In these schools, learning achievement data are used moderately, mainly to define benchmarks and targets in core subjects and competences, but is not central to schools' evaluations of their teaching and planned educational improvement. Of the teachers in this category, 62% affirmed that SIMCE results do not provide useful guidance for improving teaching or that they do so only moderately (see Table a5).

The schools in this category are reactive to PBA, but instead of following the administrative expectations *verbatim*, they adopt partial practices to fabricate better results and, in a way, 'game' the system. They adopt many of these practices reluctantly, and are aware of the limitations and risks of doing so. To start with, fabrication schools are prone to focusing their teaching on the basic competences at the center of the accountability system – so-called 'narrowing' of the curriculum. Nonetheless, despite the managers of these schools dedicating more teaching hours to the subjects externally assessed, with the clear intention of improving test results, they are also aware of the risks of these types of tactics and try to minimize them:

I also have to [ensure] that this does not go to the other extreme, because in this... search to get positive scores, other subjects are also [cast] aside [like] music [and] art; then I give more literacy when I [should really have been doing] art. Or I do more math when I [should have been doing] music. I have to be... attentive that this does not happen. (pedagogical coordinator, school 15)

Teaching to the test is, as could not be otherwise, another very frequent practice to boost test scores – with more than 50% of teachers reporting that they conduct this practice throughout the whole year, and more than 20% doing it just before delivering the SIMCE test (see Table a4). In these schools, test preparation materializes in numerous test simulations, in order to familiarize students with predictable test questions and the test format. As happens with narrowing the curriculum, the managers of these schools promote teaching to the test reluctantly and mainly in the grades covered by the SIMCE to try and minimize the risks. A common complaint, for instance, is the fact that 'instead of generating a pedagogical practice', standardized tests such as the SIMCE oblige them to engage in 'repetition of exercises to practice an evaluation style' (principal, school 13).

This is the only school category where we found that students' tracking was conducted with the explicit intention of boosting test scores, which is indicative of a fabrication response. When asked what they do to improve test scores, this principal stated:

Well, the teachers have been very creative [a]nd... have developed several strategies. One of the strategies that has given us the best result is dividing the children into groups. For example, there are three third-grade courses. So, based on the results, we divide the children. Those who achieved the learning objective, those who are average, and those who are low. [At certain times], the course structure is broken [into these groups] and the teacher who has greater competencies, because she has to have greater competencies, is the one that takes the low-performing students [...] This is what has been giving us the best results. (principal, school 14)

Tracking students seems paradoxical in a school that defines itself as 'inclusive', 'integral' and 'non-academicist' (principal, school 14). However, tracking is not a constitutive characteristic of schools' educational provision or philosophy but an experimental and pragmatic reaction to performative pressure, with the sole intention of making schools conform to performance expectations. Furthermore, as the principal of school 14 emphasized, tracking is only conducted in mathematics and literacy.

Conclusions

Schools do not respond to accountability regulations uniformly. This chapter shows that PBA regulations generate a broad and varying range of school responses, which go beyond conventional classifications mainly focusing on alignment vs decoupling or high-low fidelity implementation. Based on a mixed methods study conducted in Chile, we identified five school responses to PBA, namely accommodation, induced alignment, dilution, fabrication and *de facto* opting out. Our study does not imply that these are all possible or existing responses to accountability regulations. For instance, open and covert forms of resistance to PBA – which have not been identified in this study – are to be expected in a country like Chile, where quality assurance and accountability reforms have generated important debates, controversies and even public boycotting campaigns.

Our study highlights the mediating role of subjective variables in the enactment of PBA. Although interpretation practices have been considered important in previous enactment research, our study emphasizes the mediating role of perceived pressure in terms of

understanding how and to what extent schools make sense of and react to external accountability prerogatives. Theoretically speaking, our research shows that the way policy actors translate and respond to policy mandates is not only contingent to interpretation dynamics, but also to the levels of pressure that actors experience to comply with the policies in question. We have tested this idea by developing a unique heuristic approach in which school actors' perception of accountability pressures and their conception of the fairness and validity of PBA have allowed us to build school categories, as a preliminary step to identify schools' varying responses.

This study perspective revealed that the levels of perceived pressure among school actors in terms of PBA systems vary significantly, and they do so partly independently of objectively defined pressure measures enforced by the educational authorities. The perceived pressure factor contributes to better understanding the non-linear and often 'unexpected' nature of school responses to PBA. The school responses we identified reveal that high-stakes accountability does not operate as a performativity device in all circumstances. Not all schools that authorities subject to the 'pressure cooker' experience high levels of performative pressure. Feeling powerless in the face of quality assurance expectations, being more inclined to please audiences other than educational authorities or embracing an academicist ethos are very different factors in nature, but all they help to take the pressure off schools.

We have also identified many schools where the staff experience strong performative pressure. This includes, counterintuitively, schools with satisfactory levels of performance; however, our data reveal that performative pressure is especially strong within low-performing public schools with disadvantaged student populations. Teachers and principals in these schools tend to be critical of PBA regulations, and are those who more frequently resort to instrumental practices in an attempt to fabricate better scores and escape performance pressure. The fact that fabrication responses (and also opting out) predominate within schools serving disadvantaged students challenges to a great extent the equity rhetoric of many accountability reforms.

Overall, when accountability pressures generate high levels of instrumental and fragmented responses (as in Chile), this does not necessarily improve students' learning experiences. Indeed, alternative forms of accountability (process-based and/or oriented towards rewarding meaningful change) would be more relevant for schools in the long term and less distracting.

However, advancing these alternatives in countries like Chile is likely to be challenging as numerous accountability instruments, agents and expectations overlap (not always formally and not harmoniously), and distrust, threats and sanctions have been sedimented in the educational system reform after reform.

References

- Agencia de Calidad de la Educación (2018). *Categorías de desempeño*. Santiago, Chile. Retrieved from: <https://www.agenciaorienta.cl/>
- Agrey, L. (2004). The Pressure Cooker in Education: Standardized Assessment and High-Stakes. *Canadian Social Studies*, 38(3), 1-12.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational researcher*, 36(5), 258-267.
- Ball, S.J. (2003). The teacher's soul and the terrors of performativity. *Journal of education policy*, 18(2), 215-228.
- Ball, S.J., Maguire, M., & Braun, A. (2012). *How schools do policy: Policy enactments in secondary schools*. New York: Routledge.
- Bellei, C. & Vanni, X. (2015). The Evolution of Educational Policy in Chile 1980-2014. In: S. Schwartzman (Ed.), *Education in South America. London* (pp. 179-200). London: Bloomsbury Academic.
- Berends, M. (2015). Sociology and school choice: What we know after two decades of charter schools. *Annual Review of Sociology*, 41, 159-180.
- Boxenbaum, E. and Jonsson, S. (2017). Isomorphism, Diffusion and Decoupling: Concept Evolution and Theoretical Challenges. In: R. Greenwood; C. Oliver; T.B. Lawrence; R.E. Meyer. (eds). *The SAGE Handbook of Organizational Institutionalism* (pp. 77-101). London: Sage.
- Candido, H.H.D. (2020). Datafication in schools: enactments of quality assurance and evaluation policies in Brazil. *International Studies in Sociology of Education*, 29(1-2), 126-157.
- Coburn, C.E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of education*, 77(3), 211-244.
- Diamond, J.B. (2012). Accountability policy, school organization, and classroom practice: Partial recoupling and educational opportunity. *Education and Urban Society*, 44(2), 151-182.

- Falabella, A. (2020). The ethics of competition: accountability policy enactment in Chilean schools' everyday life. *Journal of Education Policy*, 35(1), 23-45.
- Ferrer-Esteban, G. (2020). *Sampling strategy*. REFORMED Methodological Papers No. 3.
- Jennings, J.L. (2010), School Choice or Schools' Choice? Managing in an Era of Accountability, *Sociology of Education*, 83(3), 227-247.
- Keddie, A. (2013) Thriving amid the performative demands of the contemporary audit culture: a matter of school context, *Journal of Education Policy* 28(6), 750-766.
- Kim, J. & Sunderman, G. (2005). Measuring Academic Proficiency Under the No Child Left Behind Act: Implications for Educational Equity. *Educational Researcher*, 34(8), 3-13.
- Landri, P. (2018). *Digital Governance of Education: Technology, standards and Europeanization of education*. London: Bloomsbury Publishing.
- Levatino, A. (2019). *Surveying Principals and Teachers in the framework of the REFORMED Project: Methodological Insights into the Design of the Questionnaires*. REFORMED Methodological Papers No. 2.
- Malen, B. (2006). Revisiting policy implementation as a political phenomenon: The case of reconstitution policies. In: Meredith I. Honig (Ed.). *New directions in education policy implementation: Confronting complexity* (pp. 83–104). Albany: State University of New York Press.
- Maroy, C., & Pons, X. (2019). *Accountability Policies in Education. A comparative and multilevel analysis in France and Quebec*. Cham: Springer.
- MINEDUC. (2019). *Prioridades 2019 con IVE, SINAE, BÁSICA, MEDIA Y COMUNAL*. Santiago: Ministry of Education. Retrieved from: <https://www.junaeb.cl/ive>
- Mintrop, H. (2004). *Schools on probation: How accountability works (and doesn't work)*. New York: Teachers College Press.
- Mittleman, J., & Jennings, J.L. (2018). Accountability, Achievement, and Inequality in American Public Schools: A Review of the Literature. In: Barbara Schneider (Ed.). *Handbook of the Sociology of Education in the 21st Century* (pp. 475-492). Cham: Springer.
- Mizala, A., & Schneider, B.R. (2014). Negotiating Education Reform: Teacher Evaluations and Incentives in Chile (1990–2010). *Governance*, 27(1), 87-109.
- Parcerisa, L., & Falabella, A. (2017). La consolidación del Estado evaluador a través de políticas de rendición de cuentas: trayectoria, producción y tensiones en el sistema educativo chileno. *Education Policy Analysis Archives*, (25), 1-24.

- Perryman, J., Ball, S., Maguire, M., & Braun, A. (2011). Life in the pressure cooker—School league tables and English and mathematics teachers' responses to accountability in a results-driven era. *British Journal of Educational Studies*, 59(2), 179-195.
- Skedsmo, G. (2011). Formulation and realisation of evaluation policy: inconsistencies and problematic issues. *Educational Assessment, Evaluation and Accountability*, 23(1), 5-20.
- Spillane, J.P., Diamond, J.B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy*, 16(5), 731-762.
- Tan, C. (2018). *Comparing high-performing education systems: understanding Singapore, Shanghai, and Hong Kong*. New York: Routledge.
- Teddlie, C., & Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *Research in the Schools*, 13(1), 12-28.
- Valenzuela, J.P., & Montecinos, C. (2017). Structural reforms and equity in Chilean schools, *Oxford Research Encyclopedia of Education*, 1-25.
- Watanabe, M. (2008). Tracking in the era of high-stakes state accountability reform: Case studies of classroom instruction in North Carolina. *Teachers College Record*, 110(3), 489-534.
- Zancajo, A. (2020). Schools in the Marketplace: Analysis of School Supply Responses in the Chilean Education Market. *Educational Policy*, 34(1), 43-64.

Appendix

Table a1. Share of schools by performative pressure and PBA approach, and ownership

<i>School categories</i>	Ownership			
	<i>Public</i>	<i>Private-subs.</i>	<i>Private</i>	<i>Total</i>
High pressure and con-PBA	62.5	37.5	0.0	100
High pressure and pro-PBA	33.3	66.7	0.0	100
Low pressure and con-PBA	38.9	55.6	5.6	100
Low pressure and pro-PBA	31.6	52.6	15.8	100

Source: Reformed database. Chile

Table a2. Share of schools by performative pressure and PBA approach, and socioeconomic level

<i>School categories</i>	SES quarters				Aggreg. SES quarters		Total
	Low	Low-med	Med-high	High	Low / low-med	Med-high / High	
High pressure and con-PBA	45.8	45.8	4.2	4.2	91.6	8.4	100
High pressure and pro-PBA	16.7	33.3	38.9	11.1	50.0	50.0	100
Low pressure and con-PBA	22.2	27.8	22.2	27.8	50.0	50.0	100
Low pressure and pro-PBA	26.3	10.5	21.1	42.1	36.8	63.2	100

Source: Reformed database. Chile

Table a3. Share of schools by performative pressure and PBA approach, and performance levels

<i>School categories</i>	Performance levels				Aggreg. levels		Total
	Low	Low-med	Med-high	High	Low / low-med	Med-high / High	
High pressure and con-PBA	33.3	29.2	37.5	0.0	62.5	37.5	100
High pressure and pro-PBA	5.6	27.8	55.6	11.1	33.3	66.7	100
Low pressure and con-PBA	5.6	33.3	50.0	11.1	38.9	61.1	100
Low pressure and pro-PBA	5.3	10.5	36.8	47.4	15.8	84.2	100

Source: Reformed database. Chile

Table a4. School categories and teaching to the test

<i>School categories</i>	Activities focused on preparing students for the national tests			
	<i>Never</i>	<i>Month before</i>	<i>Whole year</i>	<i>Total</i>
High pressure and con-PBA	26.81	20.65	52.54	100
High pressure and pro-PBA	25.57	13.24	61.19	100
Low pressure and con-PBA	50.23	17.67	32.09	100
Low pressure and pro-PBA	29.41	15.29	55.29	100
<i>Total</i>	32.44	16.89	50.67	100

Source: Reformed database. Chile

Table a5. School categories and data use (to improve teaching quality)

<i>School categories</i>	National test results provide useful information and guidance to improve the quality of teaching in the school			
	<i>Not at all / a little</i>	<i>Some</i>	<i>Much / completely</i>	<i>Total</i>
High pressure and con-PBA	36.0	26.1	37.9	100
High pressure and pro-PBA	21.5	30.1	48.4	100
Low pressure and con-PBA	48.5	22.6	28.9	100
Low pressure and pro-PBA	20.1	27.9	51.9	100
<i>Total</i>	31.2	26.7	42.1	100

Source: Reformed database. Chile

Table a6. School categories and data use (to assess teachers' work)

<i>School categories</i>	Uses of national test: to assess teachers' work		
	<i>No</i>	<i>Yes</i>	<i>Total</i>
High pressure and con-PBA	58.7	41.3	100
High pressure and pro-PBA	45.4	54.6	100
Low pressure and con-PBA	62.7	37.4	100
Low pressure and pro-PBA	44.8	55.2	100
<i>Total</i>	52.8	47.2	100

Source: Reformed database. Chile

Table a7. School categories and data use (to compare performance)

<i>School categories</i>	Uses of national test: to compare performance with other schools		
	<i>No</i>	<i>Yes</i>	<i>Total</i>
High pressure and con-PBA	38.8	61.2	100
High pressure and pro-PBA	32.8	67.2	100
Low pressure and con-PBA	49.4	50.6	100
Low pressure and pro-PBA	29.6	70.4	100
<i>Total</i>	37.4	62.7	100

Source: Reformed database. Chile

Table a8. School categories and data use (to build reputation)

<i>School categories</i>	Uses of national test: to build reputation		
	<i>No</i>	<i>Yes</i>	<i>Total</i>
High pressure and con-PBA	59.9	40.1	100
High pressure and pro-PBA	52.7	47.3	100
Low pressure and con-PBA	73.5	26.5	100
Low pressure and pro-PBA	52.9	47.1	100
<i>Total</i>	59.4	40.6	100

Source: Reformed database. Chile

Publication E: Parcerisa, L. (2020, under review). To align or not to align: The enactment of accountability and data-use in disadvantaged school contexts. *Educational Assessment, Evaluation and Accountability*.

CHAPTER 5

To align or not to align: The enactment of accountability and data-use in disadvantaged school contexts

Introduction

In recent years, performance-based accountability (PBA) policies have been disseminated globally as a ‘best practice’ by different international organizations such as the World Bank and the Organization for Economic Co-operation and Development (Lingard, Martino, Rezai-Rashti and Sellar 2015; Verger and Parcerisa 2018). PBA instruments are being adopted in numerous educational settings around the world with the intention of modernizing governance, strengthening school effectiveness and reducing learning gaps among different social groups. In Chile, PBA instruments were implemented in the context of a profound market-driven educational reform, introduced in the 1980s (Diaz Rios 2018; Gysling 2016; Verger, Fontdevila and Parcerisa 2019) but currently PBA generates political consensus among right-wing and center-left parties (Parcerisa and Falabella 2017). Recent educational reforms adopted by both right-wing and center-left governments have only reinforced the PBA logic to educational governance and intensified the number of national assessments, the areas covered by these assessments and the consequences of assessment results for school actors (Carrasco, Seppänen, Rinne and Falabella 2015; Falabella 2020). Nonetheless, despite the bipartisan support for PBA, important educational stakeholders are very critical of the current PBA approach and advocate the construction of a new assessment system that is more context-sensitive and formative in nature (Campos-Martínez and Fernández 2014; Flórez Petour and Rozas Assael 2019).

PBA policies follow a principal-agent logic and are not only expected to trigger school improvement through incentives and sanctions, but by promoting the use of learning-achievement data among school actors. Nonetheless, despite the expectations of PBA policies being high, especially in the context of poverty, where most underperforming schools tend to be found (Mintrop 2004), international evidence on the enactment of PBA policies and datafication shows a wide range of school responses (Falabella 2014; Koretz 2017; Landri 2018; Maroy and Pons 2019) that ranges from policy appropriation to rejection (Candido 2019) or

opting out (Landri 2018). Existing research suggests that while some schools react to PBA by adopting substantive changes to pedagogical approaches and practices, others tend to adopt cosmetic or instrumental behaviors to improve their performance in national assessments (Mintrop and Trujillo 2007a, 2007b; Mittleman and Jennings 2018). In Australia, Hardy (2015) identifies negotiations and ambivalent responses to PBA and datafication. Specifically, this investigation shows that while teachers are involved in data production, “they also seek to challenge the more reductive effects of such data generation” (Hardy 2015, p. 483).

In this sense, investigations carried out in the context of the United States indicate that educators made sense of PBA policies in complex ways, which may be complementary or even contradictory to the policy. These investigations show the relevance of cognition and meaning-making processes for understanding both school and individual responses and practices (Coburn 2005; Russell and Bray 2013; Spillane 2004). Principals’ and teachers’ perceptions in relation to PBA might be influenced by their local context, as well as ‘collective sense-making processes’ (Diamond 2012, p. 173). In this regard, an investigation carried out by Mintrop (2004) into ‘on probation’, disadvantaged schools show that negative opinions and beliefs relating to the fairness of the accountability system, predominate markedly among teachers. At school and classroom level, PBA instruments can lead to both constructive (e.g., changes in the curriculum and teaching methods, see Koretz 2008 and Hamilton, Stecher and Klein, 2002) and perverse responses (Koretz 2008; Verger and Parcerisa 2017) such as cream-skimming (Jennings 2010), teaching to the test (Au 2007) curriculum narrowing (Berliner 2011) and cheating and exclusion (Hofflinger and von Hippel 2020). In turn, perverse responses can undermine educational inclusion and quality (Falabella 2014; Lipman 2004; Thiel and Bellman 2017). According to Diamond and Spillane (2004), in the most challenging contexts, short-term and instrumental strategies are more likely to be adopted than a substantive transformation of instruction. However, based on research carried out in England, Lupton (2004) points out a variation in school responses and approaches, even in schools that work in apparently similar poverty contexts.

In Latin American countries, there is also a growing corpus of investigations that explore the enactment of datafication and PBA in disadvantaged contexts. For example, in the case of Brazil, Candido (2019) found that datafication has penetrated “the imaginaries of school actors” and

that school actors' engagement with PBA is higher and becomes more meaningful "when connected to problems individuals aim to solve" (p.14). Nevertheless, this investigation also illustrates the contentious nature of policy enactment and shows how school principals and teachers negotiate, decouple and even resist PBA policies. In Chilean disadvantaged schools, Campos-Martínez and Guerrero (2016) note that PBA policies contribute to increasing stress and trigger feelings of frustration among teachers. In terms of schools' logics of action, Falabella and Opazo (2014) identified four different school responses that include expressive, tactical and instrumental strategies. Not surprisingly, school staff from 'on probation' schools set against a background of poverty, demonstrate higher levels of pressure and stress and are more likely to adopt short-term strategies such as teaching to the test. Similarly, in an ethnographic investigation carried out in a marginalized school context, Armijo (2019) shows that even in schools with alternative pedagogical projects, PBA pressure alters school practices and promotes the emergence of tensions between standardization and a school's ethical and pedagogical principles of inclusion.

The present investigation is situated in the social and academic debate that PBA generates in Chile (see Acuña, Assaél, Contreras and Peralta 2014; Assaél, Acuña, Contreras and Corbalán 2014; Falabella 2016, 2020; Montecinos, Madrid, Fernández and Ahumada 2014; Weinstein, Marfán and Horn 2016) and aims to provide new knowledge about how Chilean school actors understand and respond to accountability policies, with a focus on contexts of disadvantages. Specifically, the objective of this study is twofold. It aims to understand, firstly, how principals and teachers interpret and experience the PBA system and secondly, how these interpretations, together with other variables of a subjective nature, mediate the enactment of accountability policies and the use of performance data among school actors.

To achieve these objectives at a theoretical level, this research draws on sense-making (Spillane 2004; Coburn 2005) and enactment theories (Ball, Maguire and Braun 2012) and, at a methodological level, the research adopts a mixed-methods design approach through which two extreme school cases (i.e., schools that perceive very different levels of performative pressure and have very different perceptions about PBA) are compared. The article is structured as follows. Firstly, I present the theoretical framework. Secondly, I describe the main features of

the current PBA system in Chile. Thirdly, I present and justify both the methods and the sample of schools. Fourthly, the two school cases are presented and finally, I discuss the main results.

Theorizing schools' responses to high-stakes accountability: meaning-making and policy translations in context

As shown above, the PBA policies can trigger a wide-range of responses and effects. To capture schools' responses to PBA, and to what extent schools align or not to PBA expectations, this article combines the policy enactment framework with sense-making theory. The framework of *policy enactment* has been developed as a critique of conventional approaches to implementation research (Ball et al. 2011, 2012). According to these scholars, policy implementation literature does not tend to pay enough attention to the mediating role of local contingencies and the agency of policy implementers. Approaches that draw from a school effectiveness tradition tend to conceive schools as dematerialized contexts and assume that implementation is a linear process where inputs are too directly translated into outcomes (Ball et al. 2012; Thrupp and Easter 2012). More recently, some investigations have pointed out the limitations of these linear and oversimplified perspectives, recognizing the complexity of the policy process (Eppel, Turner and Wolf 2011). For these scholars, the concept of enactment is more useful as a heuristic tool, to investigate how education policies are implemented in real school settings (Braun, Ball and Maguire 2011a; Braun, Ball, Maguire and Hoskins 2011b). Sense-making theory makes a similar point to enactment theory by stating that during policy implementation “actors are making meaning, being influential, contesting, constructing responses, dealing with contradictions, attempting representations of policy” (Ball 1994, p.21).

One of the most prominent contributions of this framework refers to the analytical distinction between two interrelated stages or moments: interpretation and policy translation. Interpretation refers to meaning-making processes and the decodification and analysis of external policy mandates and messages (Ball et al. 2012; Coburn 2005; Spillane 2004), whereas policy translation refers to the moment at which “schools transform their institutional agenda into specific practices, concepts, and materials as a means of responding to the policy” (Zancajo 2020, p.49). In other words, policy translation refers to those processes that entail the local adaptation or the modification of a given policy (Steiner-Khamsi 2014). Policy translation “can be selective,

picking only those components of a policy that are deemed compatible with the local environment, beneficial in the light of local conditions, or sellable to the local community” (Schulte 2018, p. 629).

Finally, policy enactment is a process that is contingent on different contextual circumstances (Braun et al. 2011b). These include: a) situated contexts, which refers mainly to the local history of the school and to the socio-economic characteristics of the school population, b) professional contexts, including the ethos, values and beliefs of the management team and the teaching staff, together with the leadership style, c) the material contexts, referring to economic resources, staffing, school facilities and technological resources, d) external contexts related to expectations, support and pressures that come from the account holder or from the policy context at the local or state level. As noted by Braun et al. (2011b), Candido (2019), Keddie (2014) and Thrupp and Lupton (2006), contextual circumstances of a different nature are crucial in explaining differences in the way schools put new policy mandates into practice. School contexts can influence not only the degree of policy appropriation, but can also trigger differences in terms of strategies, practices and policy outcomes.

The Chilean case: main features of the performative school market

In Chile, market-driven reforms in education were initiated in the 1980s by the civic-military dictatorship (1973-1990). These reforms followed an “ideological road” and were deeply “influenced by the monetarist ideas of the Chicago School of Economics” (Verger, Fontdevila and Zancajo 2016, p. 37). The cornerstones of the performative school market created by the dictatorship are free parental school choice, together with a voucher funding scheme that follows the demand, decentralization and a national large-scale assessment (known as SIMCE¹⁰) through which school results are publicly disseminated to inform choice and exit dynamics (Bellei and Vanni 2015).

¹⁰ In 1982, the dictatorship enacted the School Performance Evaluation Test (PER, in Spanish), which evolved into the current Educational Quality Measurement System or SIMCE.

The market reform introduced key elements that would contribute to the development and consolidation of an “Evaluative state” (Neave 1998) designed to govern “at a distance” (Rose and Miller 2008, p. 205) a decentralized network of autonomous educational providers, both public and private (Author(s) 2017). The reform expected that both administrative and market accountability pressures would favor competition between schools, which in turn would improve educational quality (see Chubb and Moe, 1990). Nevertheless, the implementation of market reforms triggered multiple unexpected alterations such as high levels of school segregation and undesired behavior among schools (Corvalán, Carrasco and García-Huidobro 2016; Hofflinger and von Hippel 2020; Valenzuela, Bellei and De los Ríos 2014; Verger, Bonal and Zancajo 2016).

Far from reversing the market model during the 1990s, once democracy was reinstated, the center-left coalition maintained the model’s core components and enacted a co-payment decree (Law N° 19247), through which the government made it possible for private-subsidized schools to charge fees and make a profit (Carrasco, Gutiérrez and Flores 2017; Villalobos and Quaresma 2015). Additionally, during the first mandate of the center-left coalition, the governmental action focused on increasing fiscal expenditure and improving educational supplies (e.g., teaching materials, pedagogical support) and teachers’ working conditions (Burton 2012), especially among those schools receiving the most vulnerable students. In this period, new material and reputational consequences were added to the SIMCE test. Firstly, in 1995, SIMCE’s results were disseminated for the first time with the aim of informing parents’ choice and intensifying market pressures. Secondly, in the same year, the National System of School Performance Evaluation was enacted and, as a result, SIMCE became a fundamental part of a merit-based pay system for teachers, in which collective salary bonuses became attached to school performance (Mizala and Schneider 2014).

During the early 2000s, after a decade of steady growth in public spending on education, policymakers identified certain areas of stagnation in student learning outcomes (Bellei and Vanni 2015). For this reason, the government implemented various compensatory programs directed at low-performing schools to promote data use and school improvement (Falabella 2020). In 2008, this policy model crystallized in the Preferential School Subsidy Law (Weinstein and Villalobos 2016), through which the state gives schools an additional subsidy for each

student that they enroll from a disadvantaged context (Valenzuela and Montecinos 2017). Nevertheless, to access this subsidy, schools are classified according to their performance in standardized tests and must be even more accountable to the state. In cases of poor performance, the state can impose sanctions, which include the possibility of school closure (Elacqua, Jaimovich and Román 2019). Subsequently, the Evaluative state was improved through the adoption of new data-intensive policy instruments and tools. Specifically, in 2011, the right-wing government enacted the Education Quality Assurance System (Law n° 20529), which introduced a new institutional matrix through which the state inspects, evaluates and, if necessary, sanctions both public and private-subsidized schools (Parcerisa and Falabella 2017).

In recent years, large-scale assessments and accountability have been hot topics in both political and educational debates in Chile. Since 2000, the Ministry of Education has mandated two different “SIMCE Commissions” (in 2003 and 2015) with the purpose of assessing its design and effects (Falabella 2020). In addition, together with student protests, a new social movement called Stop SIMCE has emerged to oppose standardization and test-based accountability policies (Campos-Martínez and Guerrero 2016; Pino, Oyarzún and Salinas 2016). In short, Chile has a long trajectory of accountability reforms linked to standardized tests and can be considered a unique case to study the combined effects of market and administrative pressures in the education sector.

Methods

Methodologically, this article follows a comparative case study approach (Yin 2003) based on the comparison of two *extreme cases* (cf. Danemark et al. 2002; Gerring 2007), which have been selected on the basis of a quantitative analysis that combines administrative and survey data¹¹.

Sample, data and units of analysis

In the *quantitative stage* of the research, two different surveys were administered to principals and teachers of a representative sample of 79 elementary schools (ISCED 1 and ISCED 2) located

¹¹ The data analyzed in this research comes from the REFORMED project This is an international project that analyzes the adoption and enactment of school autonomy with accountability policies in countries like Norway, the Netherlands, Spain and Chile. For more information about the project, see www.reformedproject.eu

in the three largest urban areas of Chile: the metropolitan regions of Santiago, Valparaíso, and Concepción. The sampling strategy used to choose the schools was based on the systematic probability to size sampling method and combined explicit and implicit stratification criteria (Ferrer-Esteban 2020).

The principals' survey (n=200) was administered to the principal, the pedagogical coordinator (also known as Jefe de UTP, in Chile), and the general inspector in each school. The teachers' survey (n=1130) was administered to 20 teachers in each school (10 teachers who at the time of implementing the survey taught grades and subjects evaluated in the SIMCE and 10 teachers who taught other grades and/or subjects¹²). The questionnaires¹³ included different modules related to personal and school characteristics, school demand, perception of market and administrative pressure, opinions, perceptions and beliefs about PBA, data use, commercialization, job satisfaction, etc. (Levatino 2020). The questionnaires were administered online through the specialized software Qualtrics. During the implementation process a researcher supervised and supported the participants.

The sample for the *qualitative stage* used factorial analysis to select schools with different levels of perceived accountability pressure and alignment to PBA (see Verger, Ferrer-Esteban and Parcerisa 2020, forthcoming). The factorial analysis combined data from the survey (accountability pressure index) with secondary data from the school vulnerability index database (MINEDUC 2019). Additionally, we also used the database of schools' performance categories elaborated by the Education Quality Agency (Agencia de la Calidad 2019). In total, we obtained four quadrants from the factorial analysis, which were used to select the primary education schools in which to carry out qualitative fieldwork. The two qualitative case studies were conducted in urban schools located in popular municipalities of the Metropolitan Region of Santiago.

¹² The following grades (4th, 6th and 8th) and subjects (language, mathematics, natural sciences, social sciences) are assessed in the census-based SIMCE test. For more information, see Ministerio de Educación (2016) Plan de evaluaciones nacionales e internacionales 2016-2020. Santiago: MINEDUC. Available at: https://curriculumnacional.mineduc.cl/614/articles-34980_recurso_1.pdf

¹³ Further information about the rationale and the content of the questionnaire can be found in the questionnaire methodological note: Levatino, A. (2020). Surveying Principals and Teachers in the framework of the REFORMED Project: Methodological Insights into the Design of the Questionnaires. REFORMED Methodological Papers No. 2. www.reformedproject.eu

For the purpose of this article, I rely on data from two *extreme cases* selected on the basis of the factorial analysis¹⁴. According to Gerring (2017, p. 101), “extreme-case method selects a case because of its extreme value on an independent or dependent variable of interest”. In this regard, Danemark et al. (2002) point out that researching extreme and/or pathological cases is an appropriate strategy to explore mechanisms that are normally hidden or difficult to observe in most common cases and situations. As can be seen in Table 1, although both cases have a disadvantaged socio-economic composition (quartile 2, see column “socioeconomic status”), they differ in terms of their perception of pressure as well as their alignment regarding performative instruments and goals. Thus, the two schools are located in scenarios that are opposite in terms of pressure and ideational disposition towards PBA (low alignment with PBA/low pressure and high alignment with PBA/high pressure). School A shows a low alignment with PBA and, despite being categorized as “low-performing”, perceives low levels of accountability pressure. In contrast, School B has a medium performance but perceives high accountability pressure and shows high alignment with PBA instruments and goals.

Table 1. Schools’ characteristics

School	Ownership	Administrative pressure	Performative culture	Socioeconomic status (SES)	Performance category
School A	Public	Low	Low	2	Low performance
School B	Priv. Sub.	High	High	2	Medium performance

Source: Own elaboration based on Ferrer-Esteban (2020) and Author(s) (2020, forthcoming).

Following the selection criteria mentioned above, in each case surveys were conducted with the management team (school A, n=3 and school B, n=3) and the teachers (school A, n=13¹⁵ and School B, n=20). In total, the qualitative fieldwork involved two semi-structured interviews with

¹⁴ There is a detailed explanation of how quadrants were obtained in Ferrer-Esteban, G. (2020). *Sampling strategy*. REFORMED Methodological Papers No. 3. Bellaterra: Universitat Autònoma de Barcelona. Available at: www.reformedproject.eu. For further information, see Verger, Ferrer-Esteban and Parcerisa (2020, forthcoming).

¹⁵ Since School A has fewer than 20 teachers, the sample of participants included all members of the teaching staff.

purposely selected teachers¹⁶ and one principal from each school (and in the case of School B there was also an interview with the pedagogical coordinator). The length of each interview was between 30 and 90 minutes. Complementarily, we selected key documents from each school (the institutional project and the reports from the Quality Education Agency).

Data analysis

To analyze the quantitative data, a descriptive analysis of the responses from each school was used. The quantitative analysis enables us to illustrate the context and the general patterns of response identified in each school. Regarding the qualitative data together with the institutional documents, all the interviews were transcribed and analyzed with the ATLAS.ti version 8 software. To examine the qualitative data, qualitative content analysis was carried out (Mayring 2004). Concretely, a codebook was developed, and all data were coded by combining pre-established (theory-driven) and emerging codes that covered a wide range of key themes, such as school contexts and culture, the perception of pressure, the interpretation and translation of PBA policies, data use, etc.

Findings

School A: Low-reactivity to PBA regulations

School A is a public school with more than fifty years of history. In terms of situated context, it has an approximate enrollment of 250 students and a vulnerability index of about 90%, and welcomes students who have been expelled from other schools from the neighborhood. The school's material context is characterized by a small and very poor infrastructure. Regarding the external context, the survey results show that the reputation of the school is slightly below average and school managers describe being pressured to maintain or increase enrollment (6.3 points out of 7). In terms of performance, the school has been classified as low-performing for two consecutive years. Finally, concerning the professional context, the survey data show higher levels of cooperation between teachers (see Table 2a, in the appendix) and trust (see Table 2b) in the headteacher than observed in School B. Its pedagogical project emphasizes safety and the

¹⁶ Teachers who taught in grades and subjects evaluated in the SIMCE test, and were available at the time of the visit to the school, were selected for the qualitative fieldwork.

promotion of community participation. The school's ethos is characterized by the promotion of care and respect for the environment, human rights and dignity, and interculturality.

Principled resistance to PBA and standardized tests

In School A, teachers and principals predominantly consider the PBA system to be unfair (see Table 3a), and question its validity and usefulness (see Table 3b). Specifically, School A staff consider that measuring the quality of the school on the basis of SIMCE results is unfair (40%) or very unfair (60%). A similar response pattern can be seen when teachers and principals are asked about the fairness of comparing the SIMCE scores of schools with different socioeconomic composition (100% see the uses attached to PBA as unfair or very unfair). The headteacher of School A says that he feels “resentment” towards standardized tests and the PBA system more broadly, since these instruments mainly focus on classifying schools based on their scores. His opinion is echoed in the interviews with teachers:

(I): And what do you think about external tests like SIMCE?

(I): Look, personally, I don't like them at all, because I find that they are not a... they have no arguments to say, “this school is good, this school is more or less, this school is not good, to not say bad”. Because no, it is not the reality of all schools, so it is not a test that should be the same for all the schools. (Joaquín¹⁷, lead teacher¹⁸, School A.)

In the previous quotation, Joaquín, a teacher in charge of a course evaluated by the SIMCE test, questions the validity of the test. Considering the unequal social realities existing in the country's schools, the quality of education should be measured with different instruments.

Likewise, teachers in School A are critical of the market use of SIMCE results (i.e., publication of school scores to promote school choice). Specifically, 70% of the participants express negative opinions regarding this item. When asked about the reasons for this negative opinion, a language teacher, Violeta, points out that the SIMCE test is a “discriminatory” instrument. By publishing

¹⁷ The names of interviewees are preserved for privacy purposes. For this reason, pseudonyms are used to maintain the anonymity of the participants. The research follows the European code of conduct for research integrity and complies with the ethics review procedures from the European Research Council.

¹⁸ Also known as profesor jefe, in Spanish. In Chile, the lead teacher is in charge of a grade. The lead teacher can teach various subjects beyond their own specialism. In the case of Joaquín, he taught languages and mathematics in a grade evaluated in the SIMCE test.

results based on a limited set of indicators, the test contributes to the stigmatization of low-performing schools from marginalized contexts.

(I): What do you think about the publication of the SIMCE's results or the classification [of schools]?

(I): I don't like it, I don't agree with it because, for example, it puts this school in a disadvantaged position because parents see the SIMCE results to enroll their children. Other topics such as those that I told you about [emotional support to the students] are not published. [...]Then, of course, in that sense it puts us at a disadvantage because they say: "oh look at the SIMCE's results". Then there is a kind of discrimination and ignorance of what is the vision or mission of the school, what is the [singular] characteristic that we have, which is quite particular in relation to the other 17 schools in the municipality. (Violeta, language teacher, School A.)

In other words, according to Violeta, by putting the main focus on cognitive skills, the publication of SIMCE results does not sufficiently capture the value of a school's pedagogical projects such as in School A, whose main focus is to prevent early school leaving in a context where low levels of student engagement are observed.

Similarly, teachers in School A are critical of the validity and usefulness of the SIMCE test. Many of the teachers (80%), consider that the SIMCE test does not provide useful information about student learning and that the results of SIMCE do not adequately represent what students have learned and what they can do. Violeta points out that since she does not believe in the validity and usefulness of the SIMCE test, she does not see it as very relevant. To her, a qualitative assessment system would be more meaningful to guide the teaching practice. Teachers disagree (50% disagree, 50% strongly disagree) with the statement that a good teacher can be recognized by his/her students' results in SIMCE. Similarly, high levels of disagreement are observed regarding the statement that the content of SIMCE shows what the school priorities are (40% disagree, 30% strongly disagree). In this regard, the headteacher expresses doubts about the compatibility between standardized tests and the mission and vision of the school. To him, the test is the same for all the schools, while students' access to knowledge and experiences are very different (Victor, headteacher, School A).

Another reason that School A teachers dislike SIMCE is that they have the feeling that SIMCE data makes teachers compete with each other (Violeta, language teacher, School A). According to 40% of the respondents, SIMCE data generates competition among teachers (see Table 11).

Nonetheless, as mentioned above, despite working in an “on probation” and underperforming school, teachers and managers in School A experience low levels of performance pressure coming from the public authorities. This perception of low pressure is explained by the fact that the school staff, at both the management and teaching levels, have opted out of participating in the “competition game” promoted by the PBA system due to both professional and situated contexts. To them, “school improvement” is conceived as a long-term process in which SIMCE data are not their primary proxy for success.

Data use in context: coping with cultural and material contingencies

In School A, half of the participants use SIMCE data to identify the needs of students (50%), to report teachers' results among the teaching staff (50%), or to inform parents about the school results (50%). A relatively small percentage of teachers also use SIMCE to make decisions about professional development activities for teachers (10%) (see Table 4). Nonetheless, it is in the interviews with both the headteacher and teachers that it becomes clearer that there is a rather superficial approach to SIMCE data in School A.

(T): Here at the school everyone, from the management team to the last teacher, got together to [look at SIMCE data]. Not with the importance that we should give to it, in the sense that it is not, [...], what the SIMCE score shows is not relevant for us. I insist, I see it from the point of view that this test is misdirected, to me to say: "Do you know that a school in Las Condes, Providencia, Vitacura [which are rich areas in Santiago] obtained 10 points more than us?", [...] that is, they are light-years away from [us] in comparison. (Joaquín, lead teacher, school A.)

Joaquín adds that he does not use SIMCE data in his daily teaching practices, because he finds the information that he receives from the management team, his colleagues and his own experience much more useful and meaningful:

(T): I believe more in the management team or in the same advice from my colleagues and the experience that we have with children every day. I think that for me that is much more important than external information, [...] (Joaquín, lead teacher, School A.)

Regarding the ability to use data, 60% of the respondents point out that they have little ability to do so, while 40% of the participants point out that they have some or full ability to use data (see Table 8). In addition, almost 60% of the participants in School A consider that they do not have sufficient time to analyze SIMCE data (see Table 4). However, once again, they emphasize

that if they do not do so it is because they have other priorities. To them, school improvement aspirations are still too much conditioned by the material and situated reality in which their school is positioned:

I think that much more could be done concerning what SIMCE is looking for, which is to have excellent scores, but beyond SIMCE, we could have better results. Because for example, here we do not have access to a projector, for example, we have a room where the sun comes in, that I cannot project [the slides] well, that I do not have a blackboard, rooms where it rains. (Violeta, language teacher, School A.)

The high proportion of student transfers between schools that prevails in contexts of disadvantage is due to precarious labor and housing conditions and is another factor that challenges the fulfillment of PBA expectations for data use:

(I): [In this school] have decisions been made based on test results? I do not know if you can give me an example of any decision that has been made from the analysis of the SIMCE data.

(I): [we take] practically the minimum [number of decisions based on SIMCE], because I insist again, here, how many children arrive from pre-kindergarten or kindergarten? About 15 children, but every year you get 20 new children [enrolled], who come from different schools, from different municipalities, in third and fourth grades. (Joaquín, lead teacher, School A).

Test preparation as a symbolic practice

Test preparation is a common practice among schools that want to improve their scores in national assessments. Nonetheless, in School A, 60% of the participants (see Table 5) indicate that the principal or the management team has never recommended or instructed that teaching should be more directed towards the achievement of the learning standards that are evaluated. The same percentage of teachers indicate that they never carry out “teaching to the test” practices (see Table 6). The headteacher makes it explicit that he is not very supportive of doing test simulations to prepare for SIMCE:

(I): And do you do SIMCE's test simulations?

(H): Look, some teachers generally participate in the SIMCE test, I am not very keen on the idea of... I don't mean indoctrinating...

(I): Training?

(H): to train children based on a type of test. (Victor, headteacher, School A.)

However, at the same time, the headteacher gives teachers a high level of autonomy to decide whether or not they carry out test simulations. For instance, in the case of Violeta, the pedagogical coordinator did recommend that she do two test simulations, but mainly as a symbolic practice:

(I): And at the school level, how do you prepare students for the SIMCE?

(I): Look, particularly... This year, I have noticed, neither ... I did two SIMCE tests because they [the management team] asked me to do so for a registration issue, we need evidence that [test simulations] had been made.

(I): But this was a requirement from the Education Quality Agency or from the municipality?

(I): I don't know who is demanding it, but to me, my direct boss, my technical manager, said to me: "[Violeta], let's do one or two SIMCE tests and..." As these are new courses that I have taken, it worked for me to see what were the deficiencies that the children had in terms of content and skills, which is basically that. In that sense, it helped me to work in certain areas with the kids, but... to dedicate myself to SIMCE the whole year? no, no, no, I don't have faith in the SIMCE, I have faith in the kids (laugh). (Violeta, language teacher, School A.)

Another example of the symbolic dimension that teaching to the test acquires in School A is illustrated in the following case, in which the logic of doing test simulations is to familiarize students with the standardized test format, and how to mark it correctly:

T: Is the decision to not do SIMCE simulations yours or...?

P: Yes, it's mine [decision].

T: ... or is it school policy?

P: No, it was mine; it was mine with my PIE [School Integration Program, in English] partner [...] I think we did a test simulation, but it was [for the purpose of] marking the answers. (Violeta, language teacher, School A.)

In short, School A responds superficially to PBA pressures. Despite the fact that this school is on probation, the level of data use and teaching to the test is very low. This type of response is explained, in part, by the school's disadvantaged position in the education market, but also by a prevailing organizational culture that is critical of and resistant to PBA policies and standardized tests. Thus, the case of School A challenges some of the premises of the high-stakes accountability theory of action (see Mintrop 2004), according to which threats and sanctions would contribute to change school actors' behavior and promote performative action.

School B: Aligning to PBA expectations with fidelity

School B is a private subsidized school with nearly 30 years history. In terms of situated context, it has more than 800 students and a vulnerability index of about 90%, although according to the principal, the pupils have high levels of behavioral and cognitive engagement. Thus, despite having similar rates of vulnerability to School A, the students' disposition towards the school is completely different. Regarding the material context, the school is more spacious and the infrastructure more modern than in School A. The headteacher's office is large, well-equipped, and has a mirror that generates a panopticon effect. With regard to the external context, this school is categorized as a medium-performance school and has a reputation slightly above the average (only 17.6% of the staff consider it below the average). In relation to the professional context, survey data show lower levels of cooperation between teachers (see Table 2a) and lower levels of trust (see Table 2b) in the headteacher than those observed in School A. This could be explained partly by the employment of a more hierarchical and managerial leadership style as well as by the performative pressure that permeates the school's organizational culture. As previous investigations have noted, in (Verger, Ferrer-Esteban and Parcerisa; Zancajo, 2017). It is important to note that the owner of School B is also the proprietor of other schools. In this regard, the headteacher pointed out that the schools belonging to the same owner compete with each other for performance results. One of the key points in the institutional project is the promotion of students' social mobility. In the case of School B, the school's ethos is much more oriented towards learning metrics, performativity and datafication. Thus, its main educational theme refers to promoting social mobility in poverty contexts. The school's mission and vision highlight the focus on the achievement of quality learning for all, the evaluation of learning outcomes, and educational inclusion.

Trust in numbers: a game of love and hate towards PBA instruments

In contrast to School A, in School B I found more positive opinions, perceptions and beliefs regarding the fairness, validity and usefulness of the PBA system (see Tables 3a and 3b). In School B, the use of standardized tests for market accountability purposes has a higher acceptance among school actors than in School A (see Table 3a). Half of the respondents affirm that it is fair to publicly disseminate SIMCE's results through the media and/or internet.

Following this line of thought, the pedagogical coordinator of the school highlights the importance of publishing school scores to facilitate parental choice.

(I): And what do you think about the publication of SIMCE's results and the elaboration of rankings?

(P): I would make a distinction there because the publication of the SIMCE results I think is information that parents have to contemplate, they have to possess it. The family must have access to the real data on how the school where their son [is enrolled] is working, especially in this possibility that they still have to choose [the school]. (Marcelo, pedagogical coordinator, School B.)

Critical views of some aspects of PBA are also present in this school, especially with the threat of sanctions for low-performing schools that the PBA system in Chile is in place. According to María, the headteacher of School B, punitive uses of test scores distort the potential benefits of evaluation and might contribute to exacerbating inequalities between schools, among other undesired effects.

(H): But in others, when it [the standardized test] is used with another purpose, and with a more punitive purpose, which is what has happened in Chile that has to do with the issue of mandating in a law that if you lower this or that categorization, your school is closed, I think that it loses the focus, the end, of that validated, standardized test and ... obviously, having a function that can be noble to improve or equalize the opportunities for the different schools, on the contrary yields more tribute to this great gap that exists at least in our country, and basically, [...] it becomes a threat and from there generates fear in [school] communities, stress, pressure, and [the focus on learning] is lost within the classroom [...] and many times the students begin to be trained on how they should respond to that test. (María, headteacher, School B.)

In spite of these criticisms, both the organizational and pedagogical practices in School B are clearly aligned with the PBA mandate. To a large extent, this is related to the fact that, unlike what we observed in School A, teachers in School B consider the SIMCE test to be both valid and useful. For example, almost half of the participants disagree with the statement that the results of SIMCE do not provide useful information regarding students' learning. Also, in this school, the percentage of teachers who agree with the statement that the results of SIMCE do not adequately represent what students learn is 30 points lower than in School A. Overall, what stands out in this school is that both managers and teachers emphasize the value of SIMCE data for decision-making at the school level:

(I): What do you think about standardized tests?

(P): I think they are a very good element to measure certain issues, I think it is essential that an educational system has this type of instrument because it requires a lot of information that schools could not access to otherwise. (Marcelo, pedagogical coordinator, School B.)

Despite the fact that School B's performance is good, both its management team and its teachers experience high levels of performance pressure. The headteacher considers that this pressure is related to the importance of maintaining the good results, but also to her improvement ambitions. In her own words, the school needs to "do well in terms of maintaining the category" and "from there go up" (María, headteacher, School B). Manuel, who is a teacher in mathematics, states that he does not feel very much pressured by the management team, but by the Ministry of Education and by the school owner. Carlos, who is a language teacher, considers that performance pressure is experienced "on a yearly basis", and that this pressure is especially felt by math and language teachers.

An expansive and intensive data use

School B teachers indicate that they use SIMCE data to define and monitor the school improvement plan (93.75%), to make decisions about professional development activities for teachers (62.5%), to adjust the curriculum according to the results (56.25%), to inform parents about the school results (56.25%), and to compare their school performance with that of other schools (56.25%) (see Table 4). Carlos considers SIMCE as a complementary instrument, which is helpful for monitoring the work of the school and making pedagogical decisions:

(I): And do you find the SIMCE results useful for decision-making at school?

(I): I think that every result is useful, but it will depend a lot on where you put the focus. I think knowing the result of your school in relation to the different evaluations that were applied is essential to be able to take actions, create a work plan because if [you don't have data], you work on the basis of anything, that is, what [I don't like from] SIMCE is this massive [classification of schools] that is done. (Carlos, language teacher, School B.)

In School B, 43.75% of the respondents declare that they are reasonably capable or completely capable of using learning achievement data (see Table 8), even though, similar to School A, about 75% of the respondents consider that they do not have sufficient time to analyze the data (see Table 4). To address this challenge, School B has contracted an external test service that provides them with disaggregated data on student performance at both classroom and individual level.

According to the management team and the teachers interviewed, they find these data much richer and more useful than SIMCE data for making pedagogical decisions, as well as to monitor the school improvement plan.

(E): Are the data disaggregated at the [student] level?

(P): You can do it per student. You can; for example, the [good thing] is that since it works with the information [...] that you load up in the Excel, the idea is that you can [decide] all those who are in such a category, all those who have less performance in such questions.

(E): How often [do you apply this standardized test]?

(P): We evaluate, sorry, we evaluate three times a year, a start, the initial, the intermediate to measure progress and in November. Those are three external tests that are under the same, so to speak, line. Now, yes, that is also useful to us [for the school] improvement plans. (Marcelo, pedagogical coordinator, School B.)

Regarding the performative effects of the SIMCE test in School B, a higher level of competition among teachers is observed (see Table 11). In School B, half of the respondents declare that there is a lot of competition (31.25%) or complete competition (18.75%) between teachers because of the test, while 43.75% of teachers state that there is some level of competition between them. This quote from Manuel illustrates the performative pressure that teachers in School B experience due to the PBA system.

(I): And how does the SIMCE test affect you?

(I): How does it affect me?

(I): In your daily work.

(I): I believe that pressure does not generate an anxiety that is going to kill me, but there is pressure, there is always pressure, that is what SIMCE generates for me. There is pressure [...] also, because as I was saying before, it is not funny to see the test... that your boss may have confidence in the work that one is doing, but it is not nice to see that the students are at an elementary, adequate and insufficient level. It is not something that is seen with good eyes, of course, there are economic, work and labor consequences, but there are also consequences in the mindset that one would like to have for these things. (Manuel, maths teacher, School B.)

Routinized practices of test preparation

At School B, instrumental practices to boost performance on standardized tests are prevalent. In this school, a high routinization of test preparation practices is observed. For example, 87.5% of the respondents indicate that the principal or the management team has recommended (56.25%) or instructed (31.25%) that teaching should be more directed towards the achievement of the measurable learning standards (see Table 5). Significantly, only 6.25% of respondents state that the headteacher and management team have never recommended or insisted that students

should practice for SIMCE. In this regard, 75% of the respondents declare that they have carried out teaching to the test practices. More specifically, 75% of the participants indicate that they carry out this practice during the whole year (50%) or the month before the test (25%) (see Table 6), which shows that these instrumental practices to boost school scores are, to some extent, institutionalized and embedded within the school culture.

(H): I am sure that, if we do a series of practices that are already installed, and that we have tried to consolidate and strengthen, and that we do them as part of our daily routine in a natural way, on a day-to-day basis, etc., the results should be maintained or should improve. (María, headteacher, School B.)

With the following words, Manuel illustrates to what extent teaching to the test is part of the daily life of the school:

(I): Okay, and I was going to ask you when the SIMCE dates get closer: do you do test simulations?

(I): This year, there is the eighth-grade SIMCE, and today in eighth grade, we are working with standardized programs, almost a standardized test. And we are going to work the second semester according to a work plan, me with Mario, or Mario and me, where we are going to focus on the tests. (Manuel, maths teacher, School B.)

Nonetheless, more expressive appropriations and uses of PBA instruments have been identified in School B. For instance, as stated by Carlos when asked if he could clarify what he means by having his students prepared for the SIMCE test:

(I): I try to understand that [depending on] how I teach the student to makes sense of what he reads, I am also working for the SIMCE. In other words, I do not train the student; there are schools where they prepare students for the SIMCE. So, I don't train students for the test but there are schools where they train students for the SIMCE [...], they [students] are all trained, in the subjects [evaluated] they only work on SIMCE, SIMCE, SIMCE [...], and they [these schools] have excellent results. As I tell you, what I do is to try [to guarantee that] the children understand what they read. I work on the different types of texts, [I] cover the curricular coverage [...] that the program brings me, which is the minimum [content that] I have to deliver to the student. (Carlos, language teacher, School B.)

Concerning the influence on pedagogy and curricular narrowing (Table 7), more than half of the respondents declared that SIMCE has led to some (43.75%) or much (25%) redistribution of resources in favor of the subject areas and competences that are tested. Also, 68.75% of the respondents point out that the existence of centralized learning standards has influenced the

pedagogical approach of the school, and 56.25% of them declare that SIMCE is considered when making decisions about curricular content. When asked for an example of how test results have helped with school decisions, the pedagogical coordinator of School B states that:

Look, the evaluations, for instance, we make decisions regarding curricular coverage, how it is being met or not. So, the curricular coverage that we are usually measuring, we accompanied it with the fact that when that objective was seen, it was seen in this way. (Marcelo, pedagogical coordinator, School B.)

In summary, School B follows the PBA mandate almost to the letter. This school shows a very intensive and routinized use of achievement data and resorts to commercial education services to monitor student learning. Despite this, they are aware and critical of some of the uses of PBA policies. In this school, the combination of an organizational culture that is aligned with learning achievement goals and a high perception of accountability pressure (which also comes from the school owner), translates into the intensification of instrumental practices such as teaching to the test and narrowing the curriculum. Nevertheless, this case also shows that a priori instrumental practices can be re-created and translated by teachers who adhere to more expressive logics of action.

Conclusions

In Chile, recent educational reforms have emphasized accountability and datafication as core policy instruments to promote school improvement. The theory of action of PBA assumes that schools will react to performance pressure and will use test data to promote instructional improvement measures and to address learning inequalities between different social groups. Nonetheless, by adopting a policy enactment perspective, this study shows that PBA policies might be experienced and lived very differently by school actors, and that schools might respond to PBA pressures in ways that deviate significantly from policy expectations. Even schools that apparently operate in similar contexts activate and rework PBA instruments in diverging ways. In this final section, I would like to highlight three key, interrelated conclusions that derive from this study.

Firstly, the study contributes to gaining a better understanding of under what circumstances and in which way PBA policies work in disadvantaged contexts. The overall findings presented in

this article show that schools with similar situated contexts can interpret and translate PBA policies in diverging ways (see Lupton 2004; Lupton and Thrupp 2013; Thrupp and Lupton 2006). Actors' cognition and policy perceptions interact and are co-produced with material, cultural, social and institutional factors. By carrying out an in-depth analysis of two schools that operate within a similar disadvantaged context, this article demonstrates the importance of paying more attention to variables of a subjective nature, to better understand school responses to PBA policies. Specifically, the research shows that both the perceived accountability pressure and alignment with performative culture among school actors play a key mediating role in the enactment of PBA policies.

Secondly, the study shows that there is not necessarily a correspondence between the existence of an external threat of sanctions and the levels of perceived pressure experienced by school actors. Despite the fact that School A is 'on probation' and could be closed down due to continuous poor performance, the school staff does not experience pressure and adopts very superficial - even passive - responses to accountability regulations (including low levels of data use and test preparation). This passive response could be explained by three interrelated factors, namely the lack of trust in public authorities and the PBA system (their capacity to impose sanctions), the conflicting relationship between PBA mandates and the school staff's principles, as well as the school ethos, which interact with socio-economic and material variables. On the contrary, although School B obtains good results in the standardized test, both the teaching staff and the school leaders experience much higher levels of pressure than initially expected and overreact to accountability pressures by intensifying test simulations, resorting to external services to manage additional performance data, and so on. These findings reinforce that, to understand the effects of regulatory pressures on educational actors, it is more important to take note of how these actors experience and perceive pressure rather than considering the 'objective' pressure that, according to formal regulations, actors are expected to experience (see Verger, Ferrer-Esteban and Parcerisa 2020).

Thirdly, the study challenges some of the core premises of the action theory of PBA. PBA assumes that, especially in the context of high-stakes accountability systems, the existence of external pressures will increase the motivation of headteachers and teachers working in low-performing schools, and will promote reflexivity for instructional improvement among them. It

is usually assumed that the motivational effects, triggered by PBA pressures will favor organizational and instructional changes that, in turn, will increase the educational quality of the school. However, the case of School A shows that negative interpretations and considerations of the policy translate into superficial policy implementation.

Overall, when educational actors have a negative opinion of a policy, they will tend to dilute this policy, and/or reject or avoid implementing it (Achinstein and Ogawa 2006). These dynamics are reinforced when negative interpretations accompany low levels of perceived pressure to conform with the policy in question. This is the case in School A, where, as just shown, negative perceptions with PBA predominate, at the same time as the PBA system generates a low level of performance pressure. Accordingly, as I have shown in this investigation, School A does the minimum required to comply with the PBA mandate. As Candido (2019) points out, school actors engage with data-based governance instruments to the extent that they perceive them as significant in solving existing perceived problems and improving their teaching practices. Thus, the fact that a policy is regulated and mandated is not a sufficient condition to promote educational change. Policy instruments such as PBA need legitimacy and trust in principal, to be appropriated by school actors and enacted in constructive ways. If school actors do not believe in the validity or the fairness of the policy in question, in this case PBA, they will avoid implementing it or will do the minimum to comply with regulatory requirements. Likewise, in contexts where high perceived pressure is combined with a favorable disposition toward PBA instruments and school competition, as in the case of School B, instrumental behaviors (such as teaching to the test or narrowing the curriculum) might overtake educational practices. In this case, PBA expectations of school improvement could be undermined because an excessive emphasis on test preparation could generate impoverishment of both teaching and learning experiences. This example also shows that certain contexts encourage feelings of stress and instrumental practices (e.g., teaching to the test) to become internalized and made invisible to school actors (see also Falabella 2020).

To conclude, and in terms of policy implications, the study suggests the need to design more context-sensitive and robust assessment technologies and related accountability policies. Future research should explore which accountability policy designs can contribute to the adoption of

substantive instructional and organizational changes, in order to improve students' learning experiences and teaching methods.

References

- Achinstein, B., & Ogawa, R. (2006). (In) fidelity: What the resistance of new teachers reveals about professional principles and prescriptive educational policies. *Harvard educational review*, 76(1), 30-63.
- Acuña, F., Assaél, J., Contreras, P., & Peralta, B. (2014). La traducción de los discursos de la política educativa en la cotidianidad de dos escuelas municipales chilenas: La metáfora médica como vía de análisis. *Psicoperspectivas*, 13(1), 46-55.
- Agencia de la Calidad (2019). Categorías de desempeño. Santiago. Retrieved from: <https://www.agenciaorienta.cl/>
- Assaél, J., Acuña, F., Contreras, P., & Corbalán, F. (2014). Transformaciones en la cultura escolar en el marco de la implementación de políticas de accountability en Chile. Un estudio etnográfico en dos escuelas clasificadas en recuperación. *Estudios Pedagógicos*, 40 (2), 7-26.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational researcher*, 36(5), 258-267.
- Ball, S. J. (1994). *Education Reform: A Critical and Post-Structuralist Approach*. Buckingham: Open University Press.
- Ball, S. J., Maguire, M., Braun, A., & Hoskins, K. (2011). Policy actors: Doing policy work in schools. *Discourse: Studies in the cultural politics of education*, 32(4), 625-639.
- Ball, S. J., Maguire, M., & Braun, A. (2012). *How schools do policy: Policy enactments in secondary schools*. New York: Routledge.
- Bellei, C. & Vanni, X. (2015). The Evolution of Educational Policy in Chile 1980-2014. In: S. Schwartzman (Ed.), *Education in South America*. London (p. 179-200). London: Bloomsbury Academic.

- Braun, A., Ball, S. J., & Maguire, M. (2011a). Policy enactments in schools introduction: Towards a toolbox for theory and research. *Discourse: Studies in the Cultural Politics of Education*, 32(4), 581-583.
- Braun, A., Ball, S. J., Maguire, M., & Hoskins, K. (2011b). Taking context seriously: Towards explaining policy enactments in the secondary school. *Discourse: Studies in the cultural politics of education*, 32(4), 585-596.
- Burton, G. (2012). Hegemony and frustration: education policy making in Chile under the Concertación, 1990–2010. *Latin American Perspectives*, 39(4), 34-52.
- Campos-Martínez, J., & Guerrero, P. (2016). Efectos indeseados de la medición de la calidad educativa en Chile. La respuesta de la sociedad civil. *Cadernos Cedes*, 36(100), 355-374.
- Campos-Martínez, J., & Fernández, M. B. (2014). Resistiendo la emboscada neoliberal desde las escuelas: experiencias de organización docente contra las pruebas estandarizadas. *Revista Docencia*, 52, 72-87.
- Candido, H. H. D. (2019). Datafication in schools: enactments of quality assurance and evaluation policies in Brazil. *International Studies in Sociology of Education*, 1-32.
- Carrasco, A., Gutiérrez, G., & Flores, C. (2017). Failed regulations and school composition: Selective admission practices in Chilean primary schools. *Journal of Education Policy*, 32(5), 642–672.
- Carrasco, A., & Fromm, G. (2016). How local market pressures shape leadership practices: Evidence from Chile. *Journal of Educational Administration and History*, 48(4), 290-308.
- Carrasco, A., Seppänen, P., Rinne, R., & Falabella, A. (2015). Educational accountability policy schemes in Chile and Finland. In P. Seppänen, A. Carrasco, M. Kalalahti, R. Rinne, & H. Simola (Eds.), *Contrasting dynamics in education politics of extremes* (pp. 53–80). Rotterdam: Sense Publishers.
- Coburn, C. E. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational policy*, 19(3), 476-509.
- Corvalán, J., Carrasco, A., & García-Huidobro, J. E. (Eds.) (2016). *Mercado escolar y oportunidad educativa: libertad, diversidad y desigualdad*. Santiago: Ediciones Universidad Católica de Chile.

- Danermark, B., Ekstrom, M., Jakobsen, L., & Karlsson, J. C. (2002). *Explaining Society: Critical Realism in the Social Sciences*. London: Routledge.
- Diamond, J. B. (2012). Accountability policy, school organization, and classroom practice: Partial recoupling and educational opportunity. *Education and Urban Society*, 44(2), 151-182.
- Diaz Rios, C. (2018). The role of national paradigms in the translation of test-based accountability. *Paper presented at the 2018 Conference of the Comparative and International Education Society: Remapping Global Education. North-South Dialogue*. Mexico, March 25-29.
- Elacqua, G., Jaimovich, A., & Román, A. (2019). *The effects of accountability on the allocation of school resources: Regression discontinuity evidence from Chile* (No. IDB-WP-1074). IDB Working Paper Series, 1-36. Doi: <http://dx.doi.org/10.18235/0001979>
- Eppel, E., Turner, D., & Wolf, A. Experimentation and learning in policy implementation: implications for public management. *Working paper 11/04*. Wellington: Institute of policy studies.
- Falabella, A. (2014). The Performing School: The Effects of Market & Accountability Policies. *education policy analysis archives*, 22(70), 1-29.
- Falabella, A. (2016). ¿Qué aseguran las políticas de aseguramiento de la calidad?: un estudio de casos en distintos contextos escolares. *Estudios Pedagógicos (Valdivia)*, 42(1), 107-126.
- Falabella, A. (2020). The ethics of competition: accountability policy enactment in Chilean schools' everyday life. *Journal of Education Policy*, 35(1), 23-45.
- Falabella, A., & Opazo, C. (2014). *Sistema de Aseguramiento de la Calidad y procesos de mejoramiento: una mirada desde la gestión educativa*. Informe Ejecutivo. Santiago: CIDE.
- Ferrer-Esteban, G. (2020). *Sampling strategy*. REFORMED Methodological Papers No. 3. Bellaterra: Universitat Autònoma de Barcelona.
- Flórez Petour, M. T., & Rozas Assael, T. (2019). Accountability from a social justice perspective: Criticism and proposals. *Journal of Educational Change*, 1-26. <https://doi.org/10.1007/s10833-019-09361-3>.
- Gerring, J. (2007). *Case study research: Principles and practices*. New York: Cambridge university press.

- Gysling, J. (2016). The historical development of educational assessment in Chile: 1810–2014. *Assessment in Education: Principles, Policy & Practice*, 23(1), 8-25.
- Hamilton, L. S., Stecher, B. M., & Klein, S. P. (2002). *Making sense of test-based accountability in education*. Edited by RAND. Rand Corporation.
- Hardy, I. (2015). Data, numbers and accountability: The complexity, nature and effects of data use in schools. *British Journal of Educational Studies*, 63(4), 467-486.
- Hofflinger, A., & von Hippel, P. T. (2020). Missing children: how Chilean schools evaded accountability by having low-performing students miss high-stakes tests. *Educational Assessment Evaluation and Accountability*, 1-26.
- Jennings, J. L. (2010). School choice or schools' choice? Managing in an era of accountability. *Sociology of Education*, 83(3), 227-247.
- Keddie, A. (2014). 'It's like Spiderman... with great power comes great responsibility': school autonomy, school context and the audit culture. *School Leadership & Management*, 34(5), 502-517.
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Koretz, D. (2017). *The testing charade: Pretending to make schools better*. Chicago: University of Chicago Press.
- Landri, P. (2018). *Digital Governance of Education: Technology, standards and Europeanization of education*. London: Bloomsbury Publishing.
- Levatino, A. (2020). *Surveying Principals and Teachers in the framework of the REFORMED Project: Methodological Insights into the Design of the Questionnaires*. REFORMED Methodological Papers No. 2.
- Lingard, B., Martino, W., Rezai-Rashti, G., & Sellar, S. (2015). *Globalizing educational accountabilities*. New York: Routledge.
- Lipman, P., & Haines, N. (2007). From accountability to privatization and African American exclusion: Chicago's "Renaissance 2010". *Educational policy*, 21(3), 471-502.

- Lupton, R. (2004). Schools in disadvantaged areas: recognising context and raising quality. *CASE Papers*, 076. London: London School of Economics, Centre for Analysis of Social Exclusion.
- Lupton, R., & Thrupp, M. (2013). Headteachers' readings of and responses to disadvantaged contexts: evidence from English primary schools. *British Educational Research Journal*, 39(4), 769-788.
- Maroy C., Pons X. (2019). *Accountability Policies in Education*. Dordrecht: Springer.
- Mayring, P. (2004). Qualitative Content Analysis. In U. Flick, E.v. Kardorff, and I. Steinke (eds.). *A Companion to Qualitative Research* (pp. 266-269). London: SAGE.
- Montecinos, C., Madrid, R., Fernández, M. B., & Ahumada, L. (2014). A goal orientation analysis of teachers' motivations to participate in the school self-assessment processes of a quality assurance system in Chile. *Educational Assessment, Evaluation and Accountability*, 26(3), 241-261.
- MINEDUC. (2018). *Indicadores de la Educación en Chile 2010-2016*. Santiago: Centro de Estudios, Santiago: Ministerio de Educación.
- MINEDUC. (2019). *Prioridades 2019 con IVE, SINAE, BÁSICA, MEDIA Y COMUNAL*. Santiago: Ministry of Education. Retrieved from: <https://www.junaeb.cl/ive>
- Mintrop, H. (2004). *Schools on probation: How accountability works (and doesn't work)*. New York: Teachers College Press.
- Mintrop, H., & Trujillo, T. (2007a). *School Improvement under Test-Driven Accountability: A Comparison of High-and Low-Performing Middle Schools in California* (CSE Report 717). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Mintrop, H., & Trujillo, T. (2007b). *The practical relevance of accountability systems for school improvement: A descriptive analysis of California schools* (CSE Rep. No. 713). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Mittleman, J., & Jennings, J. L. (2018). Accountability, Achievement, and Inequality in American Public Schools: A Review of the Literature. In Barbara Schneider (ed.), *Handbook of the*

Sociology of Education in the 21st Century (pp. 475-492). Cham: Springer.
<https://doi.org/10.1007/978-3-319-76694-2>

- Mizala, A., & Schneider, B. R. (2014). Negotiating Education Reform: Teacher Evaluations and Incentives in Chile (1990–2010). *Governance*, 27(1), 87-109.
- Neave, G. (1998). The evaluative state reconsidered. *European Journal of Education*, 33(3), 265-284.
- Parcerisa, L., & Falabella, A. (2017). La consolidación del Estado evaluador a través de políticas de rendición de cuentas: trayectoria, producción y tensiones en el sistema educativo chileno. *Education Policy Analysis Archives*, (25), 1-24.
- Pino, M., Oyarzún, G., & Salinas, I. (2016). Crítica a la rendición de cuentas: Narrativa de resistencia al sistema de evaluación en Chile. *Cadernos Cedes*, 36(100), 337-354.
- Rose, N. & Miller, P. (2008) *Governing the Present: administering economic, social and personal life*. Cambridge: Polity Press.
- Russell, J. L., & Bray, L. E. (2013). Crafting coherence from complex policy messages: Educators' perceptions of special education and standards-based accountability policies. *education policy analysis archives*, 21(12), 1-25.
- Schulte, B. (2018). Envisioned and enacted practices: Educational policies and the 'politics of use' in schools. *Journal of Curriculum Studies*, 50(5), 624-637.
- Spillane, J. P. (2004). *Standards deviations: How local schools misunderstand policy*. Cambridge, MA: Harvard University Press.
- Steiner-Khamsi, G. (2014). Cross-national policy borrowing: Understanding reception and translation. *Asia Pacific Journal of Education*, 34(2), 153-167.
- Thiel, C., Schweizer, S., & Bellmann, J. (2017). Rethinking Side Effects of Accountability in Education: Insights from a Multiple Methods Study in Four German School Systems. *education policy analysis archives*, 25(93).
- Thrupp, M., & Easter, A. (2012). *Research, Analysis and Insight into National Standards (RAINS) Project*. Hamilton: University of Waikato.
- Thrupp, M., & Lupton, R. (2006). Taking school contexts more seriously: The social justice challenge. *British Journal of Educational Studies*, 54(3), 308-328.

- Valenzuela, J. P., Bellei, C., & Ríos, D. D. L. (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy*, 29(2), 217-241.
- Valenzuela, J.P. & Montecinos, C. (2017). Structural reforms and equity in Chilean schools, *Oxford Research Encyclopedia of Education*, 1-25.
- Verger, A., Bonal, X., & Zancajo, A. (2016). Recontextualización de políticas y (cuasi) mercados educativos. Un análisis de las dinámicas de demanda y oferta escolar en Chile. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, 24, 1-27.
- Verger, A., Ferrer-Esteban, G., & Parcerisa, L. (2020, forthcoming). In and out of the ‘pressure cooker’: Schools’ varying responses to accountability and datafication. In: S. Grek, C. Maroy & A. Verger (eds.) *World Yearbook in Education 2021: Accountability and datafication in education*. New York: Routledge.
- Verger, A., Fontdevila, C., & Parcerisa, L. (2019). Reforming governance through policy instruments: How and to what extent standards, tests and accountability in education spread worldwide. *Discourse: Studies in the Cultural Politics of Education*, 40(2), 248-270.
- Verger, A., Fontdevila, C., & Zancajo, A. (2016). *The privatization of education: A political economy of global education reform*. Teachers College Press.
- Verger, A., & Parcerisa, L. (2017). *Accountability and Education in the Post-2015 Scenario: International Trends, Enactment Dynamics and Socio-educational Effects*. Paper Commissioned for the 2017/8 Global Education Monitoring Report, Accountability in Education: Meeting Our Commitments. Paris: UNESCO GEMR. <http://unesdoc.unesco.org/images/0025/002595/259559e.pdf>.
- Verger, A., & L. Parcerisa. (2018). Test-based Accountability and the Rise of Regulatory Governance in Education: A Review of Global Drivers. In A. Wilkins, and A. Olmedo (eds.), *Education Governance and Social Theory: Interdisciplinary Approaches to Research*, (p.139–158). London: Bloomsbury.
- Villalobos, C., & Quaresma, M. L. (2015). Sistema escolar chileno: características y consecuencias de un modelo orientado al mercado. *Convergencia*, 22(69), 63-84.

- Weinstein J, Marfán J, Horn, A. (2016) School leadership challenged by double accountability towards schools. In J. Easley and P. N. Tulowitzki (eds.). *Educational accountability: International Perspectives on Challenges and Possibilities for School Leadership* (pp. 54-72). New York: Routledge.
- Weinstein, J., & C. Villalobos. (2016). Chile: La Experiencia De La Subvención Escolar Preferencial. In M. Souto Simão (ed.), *Transferencias Directas a Escuelas: Reflexiones Sobre Prácticas En América Latina* (pp. 58–101). Buenos Aires: IIPE-Unesco.
- Yin, R. K. (2003). *Case study research: design and methods*. Thousand Oaks: Sage.
- Zancajo, A. (2020). Schools in the Marketplace: Analysis of School Supply Responses in the Chilean Education Market. *Educational Policy*, 34(1), 43-64.

Appendix

Table 2a. Cooperation between teachers

School	Question	Never	Seldom	Occasionally	Frequently
School A	...discuss teaching strategies and students' learning issues with colleagues?	0.00%	0.00%	50.00%	50.00%
	...share and/or develop instructional material(s) with your colleagues?	20.00%	10.00%	10.00%	60.00%
School B	...discuss teaching strategies and students' learning issues with colleagues?	5.56%	33.33%	27.78%	33.33%
	...share and/or develop instructional material(s) with your colleagues?	11.11%	44.44%	22.22%	22.22%

Source: Reformed database

Table 2b. Trust in principals

	Question	Not at all	A little	Some	A lot	Completely
School A	To what extent do teachers in this school feel they can consult the principal/management team when they have a problem?	0%	0%	0%	10%	90%
	To what extent do the principal/management team support teachers when they need it?	0%	0%	0%	30%	70%
	To what extent is there a cooperative effort among the teaching staff in your school?	0%	10%	10%	60%	20%
School B	To what extent do teachers in this school feel they can consult the principal/management team when they have a problem?	0%	11.11%	16.67%	33.33%	38.89%
	To what extent do the principal/management team support teachers when they need it?	0%	16.67%	16.67%	27.78%	38.89%
	To what extent is there a cooperative effort among the teaching staff in your school?	0%	33.33%	38.89%	22.22%	5.56%

Source: Reformed database

Table 3a. Fairness

Question	School A				School B			
	Very unfair	Unfair	Fair	Very fair	Very unfair	Unfair	Fair	Very fair
... to measure the quality of the school based on SIMCE's results?	60%	40%	0%	0%	25%	50%	25%	0%
... to publicly disseminate SIMCE's results in the media and/or internet]?	40%	30%	30%	0%	18.75%	31.25%	50%	0%
... that schools with different characteristics are compared on SIMCE's results?	50%	50%	0%	0%	43.75%	43.75%	12.5%	0%

Source: Reformed database

Table 3b. Validity and usefulness

Question	School A					School B				
	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
The preparation for SIMCE takes too much time away from more important activities in school	0%	10%	20%	40%	30%	6.25%	18.75%	18.75%	37.50%	18.75%
The content of SIMCE tells us what the school's priorities are	30%	40%	30%	0%	0%	6.25%	37.50%	31.25%	18.75%	6.25%
The results of SIMCE do not provide useful	0%	20%	0%	30%	50%	6.25%	37.50%	31.25%	12.50%	12.50%

information on issues with student learning										
A good teacher can be recognized by his/her students' results in SIMCE	50%	50%	0%	0%	0%	25%	37.50%	31.25%	6.25%	0%
The results of SIMCE do not adequately represent what students have learned and can do	0%	10%	10%	50%	30%	0%	18.75%	31.25%	25%	25%

Source: Reformed database

Table 4. Data use

		School A	School B
Use of national test	To define and monitor our school improvement plan.	70%	93.75%
	To identify students with a need for more support and follow-up	50%	43.75%
	To assess teachers' work	20%	37.5%
	To take decisions about professional development activities for teachers	10%	62.5%
	To inform parents about the school results	50%	56.25%
	To group students (by achievement) for instructional purposes	10%	6.25%
	To reward well-performing teachers	0%	6.25%
	To compare our performance with that of other schools	16.67%	56.25%
	To adjust the curriculum according to the results	20%	56.25%
	To report teachers' results among the teaching staff	50%	50%
	To build reputation	20%	37.5%

Main difficulties to analyze test data	The interpretation of the data requires statistical competences	20%	25%%
	Data are not provided at the student level	50%	56.25%
	Data do tell me anything I did not know before	20%	37.5%
	The report is not clear	40%	31.25%
	Lack of time to analyze them	60%	30.77%
	Data/the report are not accessible	20%	18.75%
	Other, please specify:	0%	0%

Source: Reformed database

Table 5. Instructions to practice

Question	School A			School B		
	No, never	Yes, it has been recommended	Yes, it has been instructed	No, never	Yes, it has been recommended	Yes, it has been instructed
Have the principal and the management team recommended and/or instructed that teaching should be more adjusted to the achievement of the evaluable learning standards?	60%	20%	20%	12.50%	56.25%	31.25%
Have the principal and the management team recommended and/or instructed that students should practice for SIMCE?	30%	40%	30%	6.25%	62.50%	31.25%

Source: Reformed database

Table 6. Teaching to the test

	School A			School B		
Frequency of teaching to the test	No, never	Yes, but only during the month before the test	Yes, during the whole year	No, never	Yes, but only during the month before the test	Yes, during the whole year
	60%	10%	30%	25%	25%	50%

Source: Reformed database

Table 7. Curricular narrowing and influence on pedagogy

	School A					School B				
Question	Not at all	A little	Some	Much	Completely	Not at all	A little	Some	Much	Completely
In your school has SIMCE led to a redistribution in the amount of resources (time, persons, and budget) in favor of the subject areas and competences that are tested?	40%	50%	10%	0%	0%	0%	31.25%	43.75%	25%	0%
To what extent has the existence of learning standards influenced the pedagogical approach of this school?	30%	30%	30%	10%	0%	0%	12.5%	68.75%	18.75%	0%

In your school, to what extent is SIMCE taken into account when taking decisions about curricular content?	30%	40%	20%	10%	0%	0%	25%	56.25%	18.75%	0%
--	-----	-----	-----	-----	----	----	-----	--------	--------	----

Source: Reformed database

Table 8. Capacity to use data

	School A	School B
Completely	30%	12.50%
Much	0%	31.25%
Some	10%	25.00%
A little	60%	31.25%
Not at all	0%	0%

Source: Reformed database

Table 11. Competition between teachers

Do you think it is important for teachers in this school that their students outperform those of other classes in the SIMCE test?	School A	School B
Not at all	30%	0%
A little	30%	6.25%
Some	20%	43.75%
A lot	10%	31.25%
Completely	10%	18.75%
Not applicable	0%	0%

Source: Reformed database

FINDINGS AND CONCLUSIONS

SUMMARY OF MAIN RESEARCH FINDINGS: PLACING POLICY TRAJECTORIES AND ENACTMENT PROCESSES IN CONTEXT

This section provides a global overview of the dissertation and its main findings. Following the multi-scalar logic of this doctoral thesis, the synthesis is articulated in two subsections, namely policy trajectories and policy enactment.

Global policies, diverging trajectories: a political sociology of PBA reforms

Firstly, the dissertation shows that both a socio-historic approach to policy instruments and a cultural political economy approach, are powerful analytical and heuristic devices for analyzing the adoption, evolution and mutations of public policy instruments (Publications A and B). Methodologically speaking, Publication A is based on a configurative systematic literature review (SLR). This publication shows the key mediating role of meso-level variables such as politico-administrative regimes in both the adoption and evolution of PBA instruments.

Specifically, this investigation identifies and unpacks three main rationales of adoption of PBA reform that crystalize into three divergent trajectories. The first trajectory identified in this study, occurs in countries with liberal administrative regimes (e.g., the US, UK, Australia and even Chile) where NPM reforms have been adopted to commodify the public sector. Initially, in NPM early adopters (also known as NPM marketizers), PBA instruments were enacted to promote competition among schools and inform parental school choice. In these countries, PBA instruments follow an incremental evolution, adding new policy tools to the accountability system and tending to increase both the reputational and bureaucratic stakes associated with it. In general terms, these countries are characterized by a political environment in which the relevance of PBA instruments for education governance purposes, generates a broad consensus among the main political parties. In turn, in these contexts, the consolidation of PBA instruments favors, on the one hand, the emergence of new constituencies (e.g., testing industry) and, on the other hand, new grassroots movements (e.g., the opt-out movement), which promote campaigns against educational standardization.

The second trajectory identified in this study, emerges in states with a neo-Weberian institutional tradition, namely in continental European countries (such as Austria, Denmark, Germany and Norway). In neo-Weberian states, PBA instruments were adopted as a means of promoting quality standards and strengthening the control capacity of the central government in educational systems with high levels of administrative decentralization. Unlike NPM marketizers, in neo-Weberian regimes, PBA instruments tend to be associated with lower stakes. In some of these countries, PBA instruments were initially adopted to face the "learning crisis" produced by the so-called "PISA shock" (see Ochs & Phillips, 2002; Gruber, 2006). Neo-Weberian states are also characterized by a political environment, which despite the prevalence of a broad political consensus in terms of instruments' choice, tensions and disagreements exist among the most prominent political parties regarding the desirable uses of PBA. Finally, unlike the first trajectory, the reviewed literature on neo-Weberian states does not identify the emergence of new constituencies linked to the testing industry as a significant phenomenon. In addition, despite teaching unions being generally critical to PBA policies in these countries, their reluctance mainly adopts the form of covert resistance.

Finally, the third trajectory is characterized by a partial and uneven adoption and emerges in those countries with a Napoleonic administrative tradition (e.g., France, Italy, Spain and Portugal). In these countries, PBA instruments are adopted with the primary aim of increasing administrative control over teaching work and school governance. In Napoleonic states, the sedimentation of PBA instruments is more complicated than in the aforementioned politico-administrative regimes, due to the constant back and forth dynamics. In Napoleonic states, economic and political junctures also play a critical role in the implementation of PBA instruments, which tend to suffer several discontinuities. Unlike NPM marketizers and neo-Weberian states, the adoption of PBA instruments in Napoleonic countries is highly contentious and often involves open conflict between teaching unions and central government. In line with these political dynamics, the enactment of PBA instruments at school level is more often characterized by decoupling and superficial responses than by policy alignment.

While Publication A examines PBA policy trajectories on a global scale, Publication B mainly focuses on the cultural political economy of PBA reforms in a specific national setting, such as Chile, which is the country on which most of the empirical work of this thesis focuses.

Specifically, this article aims to understand the main drivers that have mediated and favored the adoption of the so-called Quality Assurance of Education (QAE) system in Chile (2006-2011), in which numerous PBA instruments are crystallized. Methodologically, the study combines documentary analysis (which includes parliamentary records, laws, government and international organization reports, presidential speeches) and 27 semi-structured interviews with key informants (e.g., policy-makers, officials from and advisers of the Ministry of Education, researchers, student's movement activists, members of think tanks and representatives of international organizations).

Regarding the main findings, Publication B sheds light on how different evolutionary mechanisms (cf. Jessop, 2010) operated in the policy change process associated with the PBA reforms enacted in Chile. This study shows that in the context of such reforms, variation was triggered by the confluence of two crises of a different nature. This confluence favored the opening of a 'window of political opportunities' (cf. Kingdon, 1995) for the introduction of changes to the PBA system, consisting of an increase in the stakes associated with SIMCE's results and strengthening control over school governance. The first crisis can be qualified as a sort of 'learning crisis'. After a decade of substantial investment and the implementation of different school improvement programs, the perception of learning outcomes' stagnation in both national and international large-scale assessments emerged. This, coupled with a growing feeling within the government of "powerlessness" to alter schools' and teachers' behavior to install effective practices in the classroom, intersected (in the year 2006) with a profound social and political crisis, triggered by massive student protests (also known as the Penguin Revolution). Thus, performance data in both ILSAs and NLSAs have been one of the drivers that have acted as a catalyst for systemic change (see Lingard and Sellar, 2013). The second crisis was triggered by the 2006 student protests, which were the result of accumulating unease regarding neoliberal policies and the school market. According to student movement diagnosis, neoliberal reforms were the cause of high levels of inequality and school segregation in the school system. The 2006 student movement not only managed to socialize a counter-hegemonic narrative in the face of the neo-liberalization of education but was also successful in placing education as a priority issue on the government's agenda (see also Zancajo, 2019).

Publication B also analyzes why, during the policy process, PBA governance mechanisms emerged as seductive and suitable policy solutions to the crises. In particular, the selection of PBA solutions may be attributed to three main factors. Firstly, some of the more ambitious policy solutions promoted by civil society and left-wing actors, which aimed at dismantling the educational market (e.g., the end of profit and student selection) were part of a new draft bill that met with strong resistance from political opposition, the Catholic Church, private school owners, the mass media and even center-right sectors of the government. The strong hostility and resistance faced by the draft bill greatly hindered the political viability of the anti-market approach of the reform and, consequently, most of the aforementioned policy solutions were finally discarded (or diluted) since the right-wing coalition had veto-power. A second factor that favored the selection of PBA governance mechanisms is related to policy borrowing dynamics. PBA instruments were considered “best practices” by IOs and their implementation had previously been observed by Chilean policy-makers and experts through “policy tourism” practices (cf. Whitty, 2012) in various “reference societies” (cf. Schriewer, 1990). Beyond the credibility of this particular policy option, PBA became a place of consensus between the center-left government and the right-wing opposition because it allowed them to accommodate and align their political preferences. Semiotically, PBA operates as an “empty vessel” (Steiner-Khamsi, 2016), so its malleability facilitated the seduction of the so-called “Chilean duopoly”. Thus, on the part of the center-left coalition, the preference toward PBA instruments can be explained, because it allows the strengthening of the central control of the state to respond to the demand for quality education for all. In the case of the right-wing coalition, PBA instruments were perceived as a potential solution to solve market failures. They believed that PBA instruments would contribute to the sophistication of the school market, reinforcing transparency in the publication of performance results and legitimizing the free choice of school policy. Last but not least, a pragmatic cutting element relating to its technical and economic feasibility (compared to other reforms, such as the end of the co-payment, which required considerably higher investment) also influenced the selection of PBA as the main reform approach and helped guarantee the support of the Ministry of Finance for the reform.

Finally, the findings of this study also point out the key role of the World Bank in enabling the enactment of the General Law of Education (also known as LGE). Moreover, evidence from

international organizations such as WB's reports also contributed to creating a bipartisan consensus around the policy design of the new QAE system.

Unpacking school responses to PBA policies in the Chilean education system: exploring the role of subjective variables in policy enactment processes

Publications C, D and E delve into the policy enactment of PBA instruments at school level. Theoretically speaking, these publications are inspired by policy enactment theory, which is combined with decoupling and sense-making theory.

Publication C is based on a scoping review of literature on the enactment and effects of PBA policies at school level. A total of 150 documents, published between the years 1995-2014 and obtained mainly from the Web of Science (WoS) and SCOPUS databases were reviewed. This research points out the diversity of design options for accountability policies and maps the effects of PBA instruments on multiple educational dimensions. The literature reviewed identifies mixed or inconclusive results regarding the benefits of PBA policies in terms of improving student learning outcomes and reducing educational inequalities, as well as concerning the effects of PBA in terms of cooperation and competition dynamics between schools and the professional autonomy of teachers. The paper also identifies effects that go beyond those expected in the theory of action of PBA, including non-desired or unexpected effects such as teaching to the test, cream-skimming, narrowing the curriculum and educational triage, among others. This publication also notes the existence of a wide range of collective and individual responses (from open conflict to consent) to PBA policies and highlights the crucial role of the socio-economic context of schools in the recontextualization of PBA policies.

Publications D and E delve into the enactment of PBA policies in the Chilean educational system. Both publications follow a sequential mixed-methods design approach through which they intend to explore the mediating role of variables of a subjective nature in the recontextualization of PBA policies. In particular, Publication D relies on a survey of school leaders (n=200) and teachers (n=1130) in a representative sample of 79 schools. This evidence is complemented by more than 50 semi-structured interviews with school leaders and teachers, carried out in 16 schools located in the Metropolitan Region of Santiago.

Among the main findings of this publication, the identification of five school responses stands out, namely: accommodation, induced alignment, fabrication, de facto opting out and dilution. Schools categorized as *accommodation* are characterized as showing high levels of alignment with PBA regulations and perceiving low levels of performative pressure. Most of the schools within this category have a medium-high or high socio-economic status (SES) and tend to enjoy an advantaged position in the local education market (LEM). In these institutions, school improvement culture and positive perceptions of SIMCE predominate. Furthermore, these schools have a broader orientation towards data use, and test preparation practices are embedded and routinized in their daily life, although they combine both expressive and instrumental logics of action.

Despite having a high alignment to PBA regulations, unlike the previous group of schools, those that fit the category of *induced alignment*, experience great external pressure to improve or maintain their performance. These schools have a medium socioeconomic status and tend to be located in an intermediate position within the LEM. In these schools, the accountability mandate is embedded in both teaching practices and the pedagogic project. In fact, test preparation practices such as teaching to the test are carried out throughout the school year and are widely applied in all grades. In these schools, data management and tracking practices are central, and learning outcomes are usually conceived as benchmarks.

Fabrication appears in schools where negative opinions and beliefs about PBA predominate but, at the same time, these schools experience high performance pressure. More than half of the schools in this category have medium-low or low performance, and most schools have a rather disadvantaged socioeconomic status. In this group of schools, the incorporation of the accountability mandate is fragmented and superficial, and PBA instruments are partially appropriated by school leaders and teachers. Unlike induced alignment response, in these schools test preparation practices are intensive but are essentially focused on SIMCE grades and subjects. Despite the importance of learning outcomes, educational projects focus on students' comprehensive development and well-being.

The last school category refers to those schools in which low levels of perceived pressure predominate and they are critical of PBA. Interestingly, in this category, we can distinguish between two differentiated logics of action, namely *de facto* opting out and policy dilution. Both logics of action are characterized by low-intensive data-use and by the low frequency of test preparation practices. Nevertheless, the school context and the rationale for adopting these practices are very different. On the one hand, *policy dilution* appears in schools with a high SES and an advantaged position in the LEM. These schools tend to have pedagogical projects mainly focused on the development of critical thinking or socio-emotional skills, and data use practices are not guided by competition purposes or purely by learning achievement. Unlike the policy dilution response, *de facto opting out* predominates in these schools, characterized by low SES and holding a disadvantaged position within the LEM. Due to its situated context, the pedagogical project in these schools emphasizes aspects such as ‘school coexistence’ and ‘security.’ In these schools, both school leaders and teachers feel powerless due to their situated context, and low levels of data use and test preparation practices are observed.

Finally, Publication E develops a comparative case study of two extreme cases which, despite having a similar SES, are characterized by different levels of alignment with PBA mandates and perceived accountability pressure. The research triangulates evidence from survey data (n = 39) and seven semi-structured interviews with school leaders and teachers, with documentary analysis. This publication carries out a more micro-sociological analysis of two schools to gain a more fine-grained understanding of the mediating role of subjective variables in the interpretation and translation of PBA policies in disadvantaged contexts.

In Publication E, I show that in marginalized contexts, the threat of sanctions associated with test results does not necessarily translate into higher levels of school staff motivation and the adoption of data-based instructional practices and substantive pedagogical changes. Rather, one of the schools in the sample (School A) adopts cosmetic and superficial pedagogical practices to respond to PBA pressures and shows low levels of data use. The adoption of such practices is explained by the predominance of an ideational refusal to engage in external evaluations and PBA policies, or to make use of PBA instruments, which according to them, promote blaming and shaming dynamics among schools and school communities. Unlike School A, in School B, there is a higher appropriation of the PBA mandate, and there is a greater trust in numbers,

despite this being far from homogeneous. In this context, the combination of an ideational alignment toward PBA and high levels of perceived pressure triggers opportunistic behaviors, which contribute to limiting teachers' autonomy and impoverishing students' learning experiences.

CONCLUSIONS

In this thesis, I set out to analyze policy trajectories and the enactment of performance-based governance instruments in education. Firstly, I show that despite the fact that NPM and data-intensive education policy instruments (Ball, Junemann and Santori 2017; Sahlberg 2016) such as NLSAs and PBA are being spread and enacted in countries globally (Kamens and Benavot 2011; Díaz Rios 2018; Rivas and Sanchez 2020; Smith 2014; Verger, Parcerisa and Fontdevila 2019), the rationales for their adoption vary from country to country. Far from producing a global convergence around a specific model of accountability, the analysis showed that this global education reform package is not monolithic. On the contrary, the trajectories through which this global agenda is recontextualized follow a certain “path dependency” (Maroy and Pons 2019; Peters, Pierre and King 2005; Steinmo, Thelen and Longstreth 1992). In other words, this dissertation identifies three distinct trajectories (in terms of instrumentation processes and the uses associated with PBA instruments) and argues that meso-level variables such as political-administrative regimes can play a key role in explaining the cross-national divergence in the rationales of adoption and the policy trajectories of PBA instruments. In this sense, the analytical frameworks used in Publications A and B, demonstrated their potential to study how global ideas are recontextualized across multiple scales, complementing existing classical approaches to policy transfer.

Secondly, the study shows that PBA reforms are not linear. Far from conceiving education reform as static, this research shows that policy instrumentation is a highly contentious process, which in turn, is highly-sensitive to changes in the context of a material or political nature (Hay 2002). However, the findings of this research also show that once PBA instruments have been adopted, it is complicated to dismantle them for two main reasons. On the one hand, PBA instruments typically advance and evolve through layering processes and the repurposing of policy tools and techniques. This makes it easier for PBA instruments to sediment into education

systems through incremental and fragmented institutional transformations. On the other hand, the thesis highlights the crucial role of lock-in mechanisms (e.g., the ‘global education race,’ new constituencies linked to the education industry and their power of persuasion to generate broad political consensus) as facilitators of the consolidation of performance-based governance instruments.

Beyond studying PBA policy trajectories in different administrative regimes and by tracing the journey of such global education policy in Chile (2000-2011), the thesis sheds light on the drivers that have enabled and mediated the adoption and recontextualization of PBA reforms in a specific national context. Specifically, I showed how material and ideational drivers, which are operating across different scales (global and national), interact and relate dialectically during the policy change process. Likewise, the Chilean example shows that PBA's semiotic malleability, is key to understanding its success as a policy solution in a highly polarized social context. Nonetheless, the thesis also reveals the importance of political junctures and economic drivers (see also Zancajo 2019) both in the selection and in the institutionalization of the reform. In the Chilean case, far from reversing market reforms, the adoption of new PBA instruments during the 2000s served to sophisticate the “performative school market.” As is argued in Publication B, PBA reforms allowed policymakers to replace “the invisible hand of the market” by “the visible hand of the Evaluative state,” which not only regulates education services but also evaluates, rewards, punishes and classifies school actors and educational establishments, based on their performance in standardized tests (Parcerisa and Falabella 2017, p. 13).

As I show in the second section of the thesis, the adoption of particular policy instruments does not necessarily imply that they are implemented homogeneously in line with previous research (Ball, Maguire and Braun 2011; Diamond and Spillane 2004; Maroy and Pons 2019; Mintrop 2004; Landri 2018). The analysis illustrates the multifaceted and contested nature of policy enactment, which is mediated by multiple variables, ranging from those related to policy design, to the school's socio-material and professional contexts.

In terms of policy enactment, the thesis makes a double contribution to the existing literature on PBA policies in the Chilean context. Firstly, on a theoretical level, the thesis operationalizes cognitive variables that filter meaning-making processes, and develops a unique heuristic tool

that enables the measurement of both variables: the perceived performative pressure and the alignment with PBA instruments and mandates. Secondly, the sequential mixed-methods design approach with a large-scale survey also makes a methodological contribution to enactment studies in the Chilean context (see Falabella 2013). The mixed-methods design approach contributes to gaining a more fine-grained understanding of how performance-based governance instruments are interpreted and work in real institutional settings.

Finally, the findings also challenge some of the core assumptions of the theory of action of PBA. This thesis not only identifies multiple responses to the same policy but also shows that PBA policies can be appropriated and translated in many different ways, even in apparently similar socio-economic contexts. Thus, I problematize certain core assumptions of the theory of action by showing that objectively higher stakes do not necessarily entail an increase in the perceived pressure experienced by schools. Consequently, in marginalized contexts where PBA policies are perceived as unfair, the threat of sanctions might not necessarily trigger a substantive pedagogical or instructional change.

Methodological implications and limitations

The articles included in this dissertation employ different methodological strategies. Publication A carries out a systematic literature review (SLR) with a focus on PBA reforms' political economy at the international level. Although this methodological strategy has many advantages (e.g., transparency, systematicity, replicability, etc.), it is also necessary to recognize some of its limitations. For example, the database used, despite being quite comprehensive, has a clear bias towards Anglo-Saxon countries. In other words, a certain over-representation of literature is observed with a focus in countries such as the US and UK. Consequently, this methodological strategy does not allow us to adequately capture the policy change processes carried out in countries where the majority of papers on the subject are published in non-indexed journals (mostly from the Global South). Nonetheless, the incorporation of evidence from non-indexed journals would increase the risk of weakening the systematic and transparent nature of the protocol. In this sense, a trade-off is observed between the robustness of the protocol and its flexibility concerning the incorporation of non-indexed journals in the database.

These limitations were also observed in the case of publication C, which is based on a scoping review of the literature. Despite using a similar protocol, unlike SLR, this methodology does not intend to generate a new theory or new categories, nor does it need to start from a specific research question, since it has exploratory purposes. Generally speaking, the scoping-review design approach is less time-consuming than SLR. Scoping-review is particularly appropriate for examining the state of the art on a particular topic and identifying the main gaps in the literature. However, with the review conducted for Paper C we also noted the absence of investigations carried out in countries from the Global South (except for Chile). Besides, there is also a predominance of research focused on high-stakes accountability models that use test results to assign and/or distribute incentives, pedagogical support, and sanctions to school actors, to the detriment of research on soft accountability systems.

Regarding publications D and E, a large number of schools from each quadrant from the factor analysis was included to capture and interpret emerging schools' logic of action in different contexts. However, it would have been ideal expanding ethnographic fieldwork in schools, but unfortunately, this was not possible due to the temporal limitations of the research. Nevertheless, to offset these constraints, the survey data allows us to capture the perceptions, beliefs, and patterns of school responses triggered in different organizational and material school contexts.

Policy implications: the need for rethinking high-stakes accountability systems

In recent years PBA policies have globalized despite there is little conclusive evidence about their effectiveness and benefits. This doctoral dissertation shows that PBA policies can trigger various school responses in different socio-economic and organizational contexts. The research also indicates that, in disadvantaged contexts, where negative views on PBA policies prevail, the performative effects of PBA pressures can be neutralized or even inhibited, while in contexts with greater alignment towards PBA, the perception of high levels of pressure can trigger unexpected or non-desired practices. As the Campbell's law predicted decades ago, "*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 processes it is intended to monitor*" (Campbell 1975, p. 35).

Although the data is essential for decision-making and is imperative for proper educational planning, it is necessary to question the auction at which point the information provided by the SIMCE test in the Chilean context makes sense to the school communities and is useful for developing school improvement processes. Obviously, accountability measures can contribute to promote the transparency and quality of public services. However, in the case of education, and especially in the context of highly segmented educational system such as the Chilean one, outcomes-based and high-stakes accountability does not fulfill its promises. According to Leithwood and Earl (2000), it is questionable:

whether a person should be held accountable for expected performances (a) that are impossible to satisfy (e.g., ensuring that all students learn to high standards), (b) that are inconsistent with the role from which they are said to derive (e.g., teachers being held responsible for students' use of illegal substances), or (c) whose assignment and satisfaction may be quite unjustified by other factors (e.g., principals being held responsible for improving the average reading scores of students in schools with highly transient student populations). In each of these three examples, it would be legitimate to hold the teacher or principal accountable for making the most productive uses of the resources available to them in an effort to move toward the goal. But this is very different from holding them accountable for actually achieving the goal. Also, it is questionable whether a person or an organization should be held solely accountable for matters involving a shared, causal responsibility. The success of students in school, for example, is a function of many factors (Leithwood and Earl 2000, p. 4-5).

In the Chilean context, the 2019 social outbreak and the COVID-19 global pandemic have opened a window of political opportunities to rethink accountability systems and the logic of standardized tests. From my point of view, it is an excellent time to try to move towards a new model of accountability system to promote both educational equity and quality. For this, it seems necessary that accountability systems also attend to inputs and processes and that incorporate mechanisms to facilitate a “structured democratic voice” and trust relationships between stakeholders (Smith and Benavot 2019). In this sense, ensuring that teachers have “pedagogical capacity,” “instructional resources,” “sustained motivation,” and “access to clear motivation” (p. 197) are necessary - although not sufficient - conditions to promote more intelligent forms of accountability. Bob Lingard, who is one of the authors who has operationalized further intelligent accountability as an alternative to current PBA systems, considers that intelligent accountability should meet the following criteria (2009, p. 14):

- recognise the reciprocal responsibilities of all actors, including governments, systems, schools, communities and parents

- acknowledge the broad purposes of schooling (beyond academic achievement)
- reject the view that improved test results [...] (on standardized tests) are indicative of improved schooling or a more socially just school system
- reject the top-down, one-way gaze upon teachers as the sole source and solution to all schooling problems
- reject a narrow construction of school league tables and a name, blame and shame approach evident in the English schooling policy context
- recognise the centrality of informed teacher judgement and quality of pedagogies to achieving better learning outcomes for all students
- recognise the need to address poverty so as to facilitate more equal educational outcomes.

In recent years, in Chile, decision-makers have skillfully incorporated many of the criticisms raised by social movements (such as *Alto al SIMCE*) and teacher unions into the PBA system. In this sense, there has been an increasing development of soft-accountability mechanisms as well as an expansion of the number of subjects and areas evaluated in the standardized test to avoid certain undesired effects such as narrowing the curriculum. Also, it is observed a growing attention in promoting reflexivity and data-use practices among school actors, in detriment of sanctions. Nevertheless, it is not clear if such changes in the accountability uses associated to SIMCE will reverse school actors' negative perceptions as well as its side-effects, given the path-dependence that PBA instruments generate and the social relations and educational practices they reify after so many years in place.

Future lines of inquiry

Future research could try to expand the qualitative fieldwork developed in each school as well to include new schools in the qualitative sample. In this sense, it would be particularly interesting to interview school owners since, according to the data from the interviews, the school's proprietor seems to have an essential role in transmitting pressure towards school actors and making decisions about hiring commercial education services.

Likewise, it would also be interesting to delve into the relationships between PBA systems and the emergence and expansion of the education industry. For example, it would be fascinating not only to analyze how public and private providers engage with the testing industry, but also to analyze how market and accountability pressure, together with school's position in the local

education market, influences the configuration of consumption patterns and the type of contracted external services.

Concerning the current context, it would also be of great interest to investigate how the external shock produced by the COVID-19 pandemic influences school actors' perceptions about the datafication and digitalization of education, as well as examining the strategies of private corporations and EdTech entrepreneurs to expand its market niche and influence public policy to promote the privatization “of” and “in” education.

In the short term, the data collected during the development of this doctoral dissertation will allow us to deepen our knowledge about how different policy designs can favor (or not) the emergence of non-desired behaviors (such as cream-skimming, cheating and/or students' exclusion) among school actors. In this regard, we will follow a mixed methods strategy combining data from survey experiments and semi-structured interviews to get a more nuanced understanding about the influence of policy designs on principals' and teachers' behaviors.

Finally, it would also be very interesting to analyze how the possible emerging tensions derived from the interaction between the new school admission system (recently implemented in the metropolitan region of Santiago) and the PBA system are experienced at the school level from a performance and datafication perspective, from both the offer and demand sides.

References

- Ball, S. J., Junemann, C., & Santori, D. (2017). *Edu.net. globalisation and education policy mobility*. Abingdon, New York: Routledge.
- Ball, S. J., Maguire, M., & Braun, A. (2011). *How schools do policy: Policy enactments in secondary schools*. New York: Routledge.
- Campbell, D. T. (1975). Assessing the impact of planned social change. In G. Lyons (Ed.), *Social research and public policies: The Dartmouth/OECD Conference*. (Chapter 1, pp 3-45). Hanover, NH: Dartmouth College, The Public Affairs Center.

- Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality?. *Teachers college record*, 106(6), 1145-1176.
- Diaz Rios, Claudia. 2018. "The role of national paradigms in the translation of test-based accountability". Paper presented at the 2018 Conference of the Comparative and International Education Society: Remapping Global Education. North-South Dialogue. Mexico, March 25-29.
- Falabella, A. (2013). *Accountability policy effects within school markets. A study in three Chilean municipalities*. Unpublished doctoral dissertation, Institute of Education, University of London.
- Hay, C. (2002). *Political analysis* (pp. 163-167). Basingstoke: Palgrave.
- Jessop, B. (2010). Cultural political economy and critical policy studies. *Critical policy studies*, 3(3-4), 336-356.
- Kamens, D. H., & Benavot, A. (2011). National, regional and international learning assessments: Trends among developing countries, 1960–2009. *Globalisation, societies and education*, 9(2), 285-300.
- Kingdon, J. W., & Stano, E. (1995). *Agendas, alternatives, and public policies* (Vol. 45, pp. 165-169). New York: HarperCollins College Publishers.
- Landri, P. (2018). *Digital Governance of Education: Technology, standards and Europeanization of education*. London: Bloomsbury Publishing.
- Leithwood, K., & Earl, L. (2000). Educational accountability effects: An international perspective. *Peabody Journal of Education*, 75(4), 1-18.
- Lingard, B. (2009). Testing times: The need for new intelligent accountabilities for schooling. *QUTU Professional Magazine*, 24(November), 13-19.
- Mintrop, H. (2004). *Schools on probation: How accountability works (and doesn't work)*. New York: Teachers College Press.
- Maroy, C., & Pons, X. (2019). *Accountability Policies in Education*. Dordrecht: Springer.

- Ochs, K., & Phillips, D. (2002). Comparative studies and 'cross-national attraction' in education: a typology for the analysis of English interest in educational policy and provision in Germany. *Educational studies*, 28(4), 325-339.
- Peters, B. G., Pierre, J., & King, D. S. (2005). The politics of path dependency: Political conflict in historical institutionalism. *The journal of politics*, 67(4), 1275-1300.
- Rivas, A., & Sanchez, B. (2020). Race to the classroom: the governance turn in Latin American education. The emerging era of accountability, control and prescribed curriculum. *Compare: A Journal of Comparative and International Education*, 1-19.
- Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In K. Mundy, A. Green, B. Lingard, & A. Verger (Eds.), *The handbook of global education policy* (pp. 128–144). West Sussex: Wiley-Blackwell.
- Schriewer, J. (1990). The Method of Comparison and the Need for Externalization: Methodological Criteria and Sociological Concepts. In J. Schriewer & B. Holmes (Eds.) *Theories and Methods in Comparative Education* (pp. 25-83). Frankfurt am Main: Peter Lang.
- Sellar, S., & Lingard, B. (2013). The OECD and global governance in education. *Journal of education policy*, 28(5), 710-725.
- Smith, W. C. (2014). The global transformation toward testing for accountability. *education policy analysis archives*, 22(116), 1-34.
- Smith, W. C., & Benavot, A. (2019). Improving accountability in education: the importance of structured democratic voice. *Asia Pacific Education Review*, 20(2), 193-205.
- Steinmo, S., Thelen, K., and Longstreth, F. (1992). *Structuring Politics: Historical Institutionalism in Comparative Analysis*. New York: Cambridge University Press.
- Steiner-Khamsi, G. (2016). New directions in policy borrowing research. *Asia Pacific Education Review*, 17(3), 381-390.
- Whitty, G. (2012). Policy Tourism and Policy Borrowing in Education: A Trans-Atlantic Case Study. In G. Steiner-Khamsi & F. Waldow (Eds.) *World Yearbook of Education* (pp. 334-370). London: Routledge.
- Zancajo, A. (2019). Drivers and Resistances to Regulation of Education Markets: The Political Economy of the Chilean Reform, Working Paper 239.

