

# RESEARCHING THE ADOPTION OF SCHOOL AUTONOMY WITH ACCOUNTABILITY REFORMS: A METHODOLOGICAL NOTE ON COUNTRY CASE STUDIES

## REFORMED Methodological Notes No. 1

Clara Fontdevila

Universitat Autònoma de Barcelona

✉ clara.fontdevila@gmail.com

This note presents the methodological approach that guided the five case studies conducted in the context of REFORMED RS1. The case studies aimed at reconstructing and analyzing the adoption of school autonomy with accountability (SAWA) arrangements in different countries, and were informed by shared data-gathering procedures and a common analytic strategy. This note gives a detailed description of the main data-collection and data-analysis tools on which the case studies relied, and makes explicit the rationale and the theoretical premises that oriented the design of such instruments. Particular attention is given to the development of the interview guide and to the coding strategy that informed the analysis of the interview data.

### Recommended citation:

Fontdevila, C. (2019). *Researching the adoption of School Autonomy with Accountability reforms: A methodological note on country case studies*. REFORMED Methodological Papers No. 1. doi: 10.5281/zenodo.3557302

## Table of contents

1. Research background . . . . .	3
2. Overview of the research strategy . . . . .	4
Data-collection procedures . . . . .	4
Analytical approach . . . . .	5
3. Theoretical foundations . . . . .	6
4. Data-collection tools: the interview guide . . . . .	7
5. Data-analysis tools: the interview codebook . . . . .	9
6. Lessons learned, possible future directions . . . . .	10
List of references . . . . .	11
Annex 1: Interview guide . . . . .	13
Annex 2: Codebook . . . . .	19

# RESEARCHING THE ADOPTION OF SCHOOL AUTONOMY WITH ACCOUNTABILITY REFORMS: A METHODOLOGICAL NOTE ON COUNTRY CASE STUDIES

## 1. Research background

The REFORMED project analyzes how and why school autonomy with accountability (SAWA) policies are being adopted and re-formulated by policy actors operating at different scales, and inquires into the institutional frameworks and policy enactment processes that explain the different effects of SAWA at the school level. The project is structured into two interconnected research strands - Research Strand 1 (RS1) focuses on the global dissemination and adoption of SAWA policies at the regulatory level, whereas Research Strand 2 (RS2) focuses on the enactment of SAWA policies at the level of practice.

This methodological note focuses on the five country case studies conducted in the context of RS1. The case studies aimed at reconstructing and analyzing the adoption of SAWA arrangements in each one of the countries that constitute the REFORMED sample – namely, Brazil, Chile, The Netherlands, Norway and Spain. More specifically, the objectives of the case studies included (a) examining the rationale(s) behind the adoption of SAWA policies; (b) understanding how contextual factors (political, administrative, economic, cultural) mediated such process; and (c) analyzing how SAWA policies evolve over time and which factors explain these transformations.

While the country case studies were conducted independently<sup>1</sup>, they relied on shared data-collection and data-analysis instruments – backed in turn by common theoretical premises and oriented by a shared set of hypotheses, which are as follows:

- H1. SAWA reforms advance through path-dependent and contingent processes of policy configuration that are markedly conditioned by politico-administrative regimes and institutional legacies.
- H2. SAWA arrangements tend to gain autonomy from their promoters and, accordingly, adopt functions and generate effects that were not initially foreseen.
  - H2'. SAWA reforms are incremental processes that tend to advance toward higher stakes.
  - H2''. SAWA arrangements tend to evolve from a more contractual-governance form to a market-oriented form.
- H3. SAWA reforms are characterized by their capacity to accommodate different political agendas.
- H4. The participation in large-scale international assessments (ILSAs) is conducive to the advancement of SAWA reforms.
  - H4'. ILSAs are used in a selective way according to the ideological preferences of different actors, and their ultimate impact is mediated by contextual variables.

This methodological note aims at providing an overview of the methodological approach upon which these case studies relied. The objective is both to give a detailed description of data-collection and data-analysis tools, and to make explicit the rationale and methodological decisions that guided the design of such instruments.

The note is structured as follows. The next section outlines the general research strategy that oriented the case studies. It includes considerations on data-collection procedures (data sources and sampling strategies) as well as on the specifics of case-centric process-tracing as the analytic strategy of choice. The following

---

<sup>1</sup> It should thus be noted that, although the five case studies relied on a common methodological and conceptual toolbox, they were not oriented at identifying cross-national regularities. In other words, the cases were not part of a comparative research design aimed at generalizing across countries.

section of the note discusses the theoretical considerations that informed the design of data-gathering and data analysis tools. The fourth section engages with the design of the interview guide, with particular attention to the alignment of the interview questions with the objectives and hypothesis guiding the research. The fifth section describes the coding strategy that oriented the analysis of the interview data. The note concludes with a brief discussion of the limitations of this set of methodological tools as identified during the course of the research, and of the strategies developed to address them.

## 2. Overview of the research strategy

### Data-collection procedures

RS1 case studies relied on semi-structured interviews as the primary source of data collection. For each of the case studies, interviews were conducted with a wide range of stakeholders, selected on the basis of their expertise and/or direct engagement with the adoption, design or implementation of the policies of interest. The selection of the interviewees relied on a combination of purposeful and snowball sampling.

*Purposeful sampling* strategies (also referred to as purposive sampling strategies) are those oriented at ensuring that certain individuals whose perspective is judged unique or relevant are included in the final sample. The selection of participants is criterion-based and relies on the a-priori knowledge of the researcher(s). Individuals are thus selected because some of the features they display (specific experiences, affiliations, etc) are judged of particular interest for the objects of the research (Mason, 2002; Patton, 2015). Thus, RS1 case studies selected participants in view of their position or role during the process of interest. In particular, individuals were selected on the basis of their influence during the process of policy formation of SAWA arrangements, or, alternatively, given their privileged role as direct observers of the process and/or their status as experts on such questions.

On the basis of this purposive sampling approach, a first list of (prospective) interviewees and informants was identified before the start of the fieldwork. As discussed in more detail below, this first sample was primarily informed by secondary sources. Later on, the sample was expanded and refined on the basis of a *snowballing approach*. Thus, when judged appropriate, interviewees were asked to identify other potentially relevant interviewees. Snowballing was essentially used as a supplementary approach, hence avoiding the risk of selection bias typically associated to this strategy (Ritchie et al., 2003). Overall, the selection process had an iterative nature – that is, the sample was revisited and refined on multiple occasions, according to the advances made in the understanding of the process of interest (the adoption of SAWA).

The final sample of interviewees included representatives from 6 key groups of particular relevance for the purposes of the research – namely, top-level politicians, policy-makers (including senior civil servants, national and local government officers directly involved in the policy design of SAWA instruments), members of government advisory committees, representatives of teachers', school leaders' and student's unions, experts (including academics and journalists) and other relevant education stakeholders (including parent organizations and staff from education authorities not directly involved in policy-shaping processes). The final breakdown is as follows:

	Brazil	Chile	Norway	Spain	The Netherlands
Top-level politicians	3	4	4	5	0
Policy-makers	3	6	7	13	7
Advisory committees	4	4	4	3	5
Teacher and student unions	5	6	11	4	2
Experts	5	4	12	4	5
Other education stakeholders	3	10	2	13	12
<b>Total</b>	<b>23</b>	<b>34</b>	<b>40</b>	<b>42</b>	<b>31</b>

It is worth noting that both purposeful and snowballing sampling strategies fall within a non-probability sampling approach. This was deemed appropriate given the ultimate object of the case studies (oriented at reconstructing the adoption of SAWA policies and gaining an understanding of which were the predisposing and precipitating factors), as well as the process-tracing approach that guided the case studies (see below).

While probability-sampling strategies are considered a necessary pre-condition for those studies oriented at generalizing their findings to a larger population, this is not so for process-tracing analysis aimed at gaining a fine-grained understanding of an event (or chain of events). Process-tracing analyses are thus primarily concerned with the incorporation of all those relevant or influential actors – rather than with representativeness and generalizability (Tansey, 2007).

It should finally be noted that, while this methodological note focuses on the strategies used in the generation and analysis of interview data, such information was supplemented with secondary data sources. These included legislative frameworks and drafts, green and white papers, minutes or summaries of parliamentary debates, research reports and studies (commissioned by public authorities but also by international organizations or private organizations), position papers and public statements of different education stakeholders, and media sources.

The use given to these secondary data varied across case studies, as it was ultimately dependent on the specificities of the SAWA policy of interest, as well as on contextual particularities. In general terms, documentary data was used for five different purposes, namely (a) understand the broader policy context in which SAWA policies were designed and adopted; (b) understand the implementation of the policies as well as their reception among the general public and specific groups of stakeholders (the teaching profession, parents, etc); (c) understand the evolution of the design and uses of SAWA policy instruments; (d) inform the conduction of interviews (identifying key themes and debates, relevant actors); and (e) contrast and complement interviewees' accounts.

## Analytical approach

In order to interpret and make sense of the collected data, RS1 case studies turned to a process-tracing approach – in particular, to the specific variant that Beach and Pedersen denominate case-centric process tracing (2016). Broadly speaking, process-tracing methods are those explicitly aimed at unearthing the causal process behind an outcome of interest, and which put a premium on the identification of causal mechanisms (George & Bennett, 2005)<sup>2</sup>. In addition to the preference for mechanism-based explanations, other distinct features of process-tracing include the reliance on fine-grained descriptions and the concern for sequencing of events, that is, a preoccupation for the unfolding of a given phenomenon and its evolution over time (Collier, 2011).

A process-tracing approach was considered particularly apposite for the purposes of RS1 case studies in that these aimed at reconstructing the chain of events eventually leading to the adoption of SAWA policies, gaining insight on the evolution of both the design and uses of SAWA policies with the passage of time, and identifying the enabling and precipitating factors that contributed to such processes.

Despite the commonalities listed above, process-tracing operates essentially as an umbrella term encompassing a wide range of approaches and accommodating a diversity of objectives (Waldner, 2012). The distinction between case-centric and theory-centric studies, advanced by Beach and Pedersen (2016), is of

---

<sup>2</sup> While the notion of causal mechanism is a notoriously contested one, an oft-cited and widely accepted definition is the one advanced by Glennan (2002), that is, “a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations” (p. S344).

particular relevance here. *Case-centric studies* are those oriented at articulating exhaustive explanations for the outcome of interest, frequently adopting an idiographic perspective, whereas theory-centered studies are oriented at testing more generalizable mechanisms. Since the goal of the RS1 case studies was to provide a comprehensive, in-depth understanding of the formation of SAWA policies, they relied on a case-centric variant of process-tracing. Case-centric approaches, also referred to as explaining-outcome process tracing, exhibit specificities with relevant methodological implications. Most notably, case-centric approaches appear to rely on a distinctly broad understanding of the notion of causal mechanism that includes case-specific mechanisms as well as conglomerates of mechanisms (as opposed to universal or context-invariant mechanisms). As a consequence, case-centric process-tracing does not lend itself to generalization or extrapolation purposes, although it may suggest certain patterns and regularities requiring further research. Another particularity of case-centric approaches is their theoretical eclecticism. Since such process-tracing variant is oriented at capturing the multiplicity of causes and processes behind a given outcome, it allows for (and even requires) the combination of different theoretical approaches (Beach & Pedersen, 2016).

### 3. Theoretical foundations

This section provides an overview of the theoretical approaches that informed RS1 case studies – and particularly, of those that informed the design of the data-collection and data-analysis tools discussed over the next sections<sup>3</sup>.

In order to capture and make sense of the adoption of SAWA policies, RS1 case studies relied primarily on the insights made by the multiple-stream approach (MSA) as advanced by Kingdon in his ground-breaking book *Agendas, Alternatives, and Public Policies* (1984). Kingdon contends that public policy-making entails at least four different processes, namely, the setting of the agenda, the specification of alternatives, the choice among the proposed alternatives, and the implementation of such decision; and argues that the first two of them (agenda-setting and alternative generation) can be explained on the basis of three (analytically distinct) streams that “flow” through the system . These are as follows:

- *The problem stream* – in which a specific issue starts receiving policy-makers attention, frequently as a result of a focusing event that directs attention to a particular source of concern, as well as of indicators and feedback mechanisms generating interest around certain conditions. Such conditions do eventually evolve into problems only if the disparity between a given situation and an ideal state of the affairs is perceived to require governmental action.
- *The policy stream* – in which a series of proposals and alternatives are developed and taken into consideration within communities of specialists. Frequently, such process entails the recombination of already-existing alternatives; and the survival of a given proposal is often the combination of a number of qualities, including its technical and political feasibility, financial workability and their fit with dominant values .
- *The political stream* – in which certain agenda items and/or alternatives gain prominence and generate consensus among key decision-makers as a result of a variety of factors including the so-called “national mood”, the balance of organized political forces, and changes within the government or administration.

---

<sup>3</sup> As noted in the previous section, case-centric studies do often require the combination of different theories in order to craft a plausible and exhaustive explanation for an outcome of interest. This section focuses on MSA given its centrality in the design of data-collection and data-analysis instruments, and as it features among the theoretical perspectives common to the different case studies. However, it should be noted that each one of the case studies combined the insights advanced by MSA with other theoretical and frameworks deemed most appropriate to make sense of the collected data. This included, most notably, insights from the Cultural Political Economy (Jessop, 2010), as well as from the policy instruments approach as advanced by Lascoumes and Le Galès (2007).

Kingdon goes on to argue that the three streams need to be brought together in order for policy change to occur. Such critical episodes are termed policy windows and owe much to the role and labor of policy entrepreneurs who actively attach problems to solutions and advance them in politically propitious situations (Kingdon, 1984)<sup>4</sup>.

MSA constitutes what John (2003) termed a *synthetic* account of policy change in that it weaves together research on different causal processes and key public-policy themes— specifically, on institutions, networks, socioeconomic processes, choices and ideas. First, it represents a promising approach to provide an exhaustive and fine-grained explanation of a given outcome as required by a case-centric approach to process tracing. A second strength attributed to MSA is its flexibility and universality – in the sense that the main concepts that configure this approach can be adapted to a variety of contexts despite Kingdon’s original focus on the US federal government (Cairney & Jones, 2016). And thirdly, MSA is particularly appropriate for the purposes of RS1 case studies in that it brings to the fore the cumulative and evolving nature of policy-formation, and particularly the incremental nature of the development of alternatives and solutions<sup>5</sup>. MSA remains thus a powerful heuristic device to identify both the immediate triggers of policy change as well as the contextual factors and gradual transformations that contributed to create a receptive environment for this to occur.

To be sure, it should be noted that theories of policy processes, including MSA, have rarely engaged in the identification of causal mechanisms as required by a process-tracing approach. However, van der Heijden, Kuhlmann, Lindquist and Wellstead (2019) have documented a growing interest for causal-mechanisms within the policy studies literature. The authors note indeed that MSA “might be the most aligned framework with causal mechanistic analysis given its focus on factors that contribute to the opening of a policy window as a trigger” (p. 8) – a point already raised by Capano (2013), who noted that the assumptions informing MSA are largely in line with causal-mechanistic thinking. Van der Heijden and collaborators (2019) go on to note that MSA tacitly identifies a number of elements likely to operate as mechanism triggers (including changes in government composition, shifts in national mood, pressure campaigns, focusing events, etc), as well as at least two crucial mechanisms – namely, the coupling of different streams as a result of policy entrepreneurs’ strategies; and the growing receptivity to ideas and experience coming from elsewhere (that is, the spillover effects from other areas, jurisdictions or policy levels).

#### 4. Data-collection tools: the interview guide

For the conduction of interviews, RS1 case studies relied on a semi-structured questionnaire format. This section gives an overview to the alignment of the interview questions with the objectives, hypothesis and dimensions of analysis guiding the research – and discussed in the previous sections.

The **first module** of the interview guide (*Starting questions*) was the simplest and shortest one, as it was oriented at gaining some basic insights on the background and expertise of the interviewee, as well as on their knowledge of the origins and recent evolution of the SAWA agenda in their own national context. This first block was thus essentially oriented at tailoring the following questions according to their familiarity of the interviewee with the object of the research, but also on refining the researchers’ understanding of the SAWA scheme in place.

The **second** and main module of the script focused on the process of *policy formation*. It grappled thus with questions relative to the adoption of SAWA policy instruments – that is, with the process through which such

---

<sup>4</sup> It should be noted that the three streams should not be understood as part of a linear or sequential model- that is, they do not constitute neat stages following one another. In fact Kingdon notes that solutions can in fact precede problems (1984; see also Béland & Howlett, 2016 for an extended discussion on the theoretical relevance of such cases).

<sup>5</sup> The incrementalism that characterizes the development of policy alternatives contrasts however with the more volatile character of the emergence of agendas, which are frequently subject to relatively dramatic or abrupt change (Kingdon, 1984).

policies and/or the problems they were expected to address gained relevance in the agenda. The block was organized into three main subsections, corresponding to the problem, policies and politics streams identified by MSA. For each one of these sub-sections, particular emphasis was put on the identification of triggers of policy change, as well as on the nature of the actors engaged in the process, their motivations, strategies and ideational sources. The module thus addressed H1 (contextual factors influencing policy configuration), H2 (evolution of SAWA regimes) and H3 (political agendas behind SAWA reforms).

A key device used in this second module was a chart summarizing the different SAWA-related policy instruments and tools in place in the country (or sub-national context) of interest. This “SAWA chart” had a double purpose. First, it contributed to make the object of the research (that is, SAWA policies) more intelligible or clearer to the interviewee. This was judged appropriate given that the existence (or meaning) of a SAWA framework, at least for certain interviewees, was not self-evident, even if they were familiar with some (or most) of its constituent policies. The chart offered an opportunity to expound and delineate the boundaries of an analytical category (SAWA) in terms that were meaningful to the interviewee.

Secondly, the SAWA chart helped the interviewer gaining an understanding of the policy arrangements the interviewee was more familiar with – thus complementing the information elicited in the first module. Interviewees were asked about the specific policy instruments they had been engaged with (whether in their design, implementation, etc) or the components of the SAWA framework they were more knowledgeable about. The interviewee’s response to such questions served to determine which instruments should be addressed in the interview, and assess whether it made sense to address them separately or whether they could be addressed as a whole (for instance, if the interviewee understood them as part of a single and coherent system).

The **third module** inquired first into the *enactment and implementation* of SAWA policies. Particular attention was paid to the evolution and impact of autonomy schemes, the uses given to standardized tests, typically lying at the core of SAWA schemes, and the role of the private sector – as a case in point of an actor whose role in the education system has often been enhanced or at least transformed by the introduction of SAWA measures (Verger et al., 2019). Although such questions were not directly related to the original set of hypotheses guiding the research, the module contributed towards a better understanding of the evolution of SAWA policies, thus addressing H2.

Additionally, the third module explored the *subjective experience or opinion* of the interviewee regarding the potentials and dangers of the SAWA agenda. This was oriented both at gaining a better understanding of the actual enactment of SAWA policies but also to understand its impact on the identities and roles of key education stakeholders. Also, gaining insight into the interviewee’s ideological alignment and genuine trust on SAWA was also considered relevant in order to better contextualize their accounts of the reform approach<sup>6</sup>.

While this subset of questions relied generally on a direct approach, an indirect or proxy question was also developed in order to determine which aspects were perceived as more problematic by the interviewee even if they did not adopt a critical stance towards SAWA. The interviewee was thus asked to reflect on which would be the characteristics of a new accountability system if the Ministry of Education ask them to (re)design it (see question 3.31 in Annex 1).

The **fourth module** was primarily oriented at identifying the *ideational sources* having influenced significantly the emergence and design of SAWA policies. The module elaborated into some of the questions possibly raised in the third module – that is, how knowledge production and knowledge dissemination dynamics had contributed to policy change, impacting each one of the three streams of policy processes. The module addressed H1 in that it explored a particular variety of drivers of policy change- namely, ideational factors. The module engaged also with H4 as particular attention was put on the role of international large-scale assessments and their influence on the processes of policy change.

---

<sup>6</sup> This was of particular relevance given the ideologically-sensitive nature of SAWA debates. Such political dimension makes it particularly ill-advised to take the interviewees’ accounts at face value. Information regarding interviewees’ attitudes and feelings towards SAWA were thus key in order to better interpret their reconstruction of the facts.

The **fifth** and final block was oriented at *wrapping up* the interview and, most importantly, asking for *other potential interviewees* – as per the snowballing sampling strategy described above.

The final interview guide can be consulted in Annex 1 of this document. As discussed above, it should however be noted that the guide was always tailored to the areas of expertise and particular experience of the interviewee in relation to SAWA, as well as to the specificities of the national context and the policy of interest. Thus, not all the questions and modules were necessarily covered for all the interviewees.

## 5. Data-analysis tools: the interview codebook

In order to analyze the data collected through the interviews, RS1 case studies relied initially on a structural-coding approach. **Structural coding** encompasses those attempts to simultaneously segment (that is, classify) and code interview data by applying a question-based approach (Saldaña, 2009). Thus, data is analyzed according to the research questions used to frame and organize the interview – as opposed to more inductive approaches as those informed by grounded theory. The strategy is oriented at facilitating the identification of every fragment associated with a theme and is considered appropriate for a variety of research designs – including hypothesis-testing approaches relying on semi-structured protocols, as in the case of RS1 case studies.

To be sure, this approach is frequently devised only as a first step, necessary to facilitate further and more in-depth analysis – both within and across the identified themes (MacQueen, McLellan, Kay, & Milstein, 1998). For the purposes of RS1 case studies, this second scanning relied on the application of **analytic codes** as defined by Deterding and Waters (2018). According to these authors “analytic codes represent the concepts to explore in a single paper or book chapter and integrate emergent findings with what is known from the literature” (p. 15). Such a coding strategy needs to be understood as part of a **flexible-coding approach** which, again, represents a departure from strategies inspired by grounded theory – in that they rely largely on theoretically-informed categories which are frequently defined a priori (although they allow for the incorporation of emergent themes identified only after the first reading and segmenting of the transcripts).

In the context of RS1 case studies, the structural coding conducted on the basis of a selected set of interview transcripts allowed for the identification of 6 main groups of codes corresponding to broadly defined topics. This included **Attributes of the interviewee**; **Subjective perception of SAWA regimes**; **Policy process: diagnosis**; **Policy process: policy formation**; **Policy process: balance of forces, actors**; and **Knowledge mobilization**. This first set of codes corresponded roughly to the 5 modules organizing the interview guide. This, in turn, echoes the hypothesis and theoretical considerations guiding the research and particularly the insights advanced by MSA – most notably, the analytic distinction between the problem stream (“diagnosis”), the policies stream (“policy formation”) and the political stream (“balance of forces”).

This first set of (macro)codes was subsequently refined and complemented with a list of analytic codes, corresponding to key concepts, actors and causal mechanisms identified by the reviewed literature. This set of analytic codes was thus common to the different case studies, although it was combined with more context-sensitive categories specific to each country or region of interest. The reliance on a shared set of codes represented however an important challenge in terms of code comparability and intercoder reliability. In order to ensure codes were used in a consistent way, and as suggested by MacQueen, McLellan, Kay and Milstein (1998), a codebook was developed. The codebook worked thus as a set of “coding guidelines”. Each code contained a series of entries oriented at setting the boundaries of the notion it intended to capture – including the name of the code, a brief and a full definition, precisions on when to use and when not to use it, one or two examples (e.g. fictitious quotes exemplifying appropriate context of use), and the group (or macro-code) to which they belong. While the complete codebook can be consulted in full at Annex 2 of the note, the following table offers an example of the typical structure of a codebook entry.

Code	POL_NEGOTIATION
<i>Brief definition</i>	Negotiation dynamics
<i>Full definition</i>	Mentions to political dynamics contributing to consensus-building and to the advance of a particular proposal, including political bargaining, compromises, concessions, etc).
<i>When not to use</i>	Does not refer to “persuasion” dynamics in which a given actor genuinely changes their preferences or priorities.
<i>Example</i>	“After two days of massive strike the government withdrew the provisions on early retirement”.
<i>Group</i>	Policy process: policy formation

Finally, it should be noted that reliance on a codebook was combined with the use of interview-specific memos oriented at organizing the information and identifying particularly interesting or relevant interviews, but also at detecting emergent topics, problematic code choices, hypothesis or future research directions. While memos were essentially open-ended in nature, each one of them included at least a definition of the dependent outcome (policy change) and the governance level addressed by the interview; the main focus of the interview and its value or utility (i.e., in which ways or to what extent it represented a significant contribution); and the knowledge and familiarity of the interviewee regarding the SAWA regime.

## 6. Lessons learned, possible future directions

This section highlights some of the main limitations of the methodological approach discussed in this note – as identified during the course of the fieldwork or during the analysis of the data corpus.

One of the main challenges encountered during the conduction of the fieldwork relates to the retrospective nature of the research. This was particularly problematic for those interviews focusing on long-established policy instruments, whose design might date back a decade or more. Eliciting information on these relatively remote events thus proved difficult in that it required considerable effort on the part of the interviewee to make the most of the interview – and also to avoid creating an uncomfortable or frustrating situation if the interviewee felt unprepared. In addition, such backward-looking accounts posed important challenges for the analysis of the data – in that the accuracy and reliability of the data was sometimes uncertain and required additional validation through secondary data.

The focus on SAWA policies as *isolated* from other system features constituted an associated difficulty. Given the analytic nature of the SAWA category, some interviewees found it difficult to focus on the design or origins of these particular policy instruments, or struggled to engage with some of the questions. While the use of the chart contributed to rendering the object of the research meaningful in the eyes of the interviewees, sometimes, some of them remained confused by the emphasis on certain policy tools that they perceived as secondary or unimportant. This was particularly the case of those interviews discussing policy instruments with a long tradition in the education system addressed by the case study. The naturalization of such instruments made it particularly challenging for some interviewees to reflect on the factors explaining their adoption or even their preservation over time.

Finally, the multi-dimensional, cumulative and complex nature of the object of study (e.g. SAWA arrangements) represented an important challenge in that it required an important selection effort on the part of the interviewer. Given the impossibility to address each one of the components of the SAWA regime (or the different forms they have adopted over time), it was necessary to tailor the interview guide according to the expertise and experience of the interviewee with the policies. The personal discernment of the researcher thus played a key role in identifying the most fruitful lines of inquiry, but also in framing the questions in a sufficiently flexible way – one that allowed the interviewee to elaborate on those areas they were more knowledgeable about.

## List of references

- Beach, D. & Pedersen, R. B. (2016). *Causal Case Study Methods. Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Arbor, MI: University of Michigan Press.
- Béland, D., & Howlett, M. (2016). How solutions chase problems: Instrument constituencies in the policy process. *Governance*, 29(3), 393–409.
- Cairney, P., & Jones, M. D. (2016). Kingdon's Multiple Streams Approach: What is the empirical impact of this universal theory? *Policy Studies Journal*, 44(1), 37–58.
- Capano, G. (2013). Policy dynamics and change: The never-ending puzzle. In E. Araral Jr., S. Fritzen, M. Howlett, M. Ramesh, & X. Wu (Eds.), *Routledge Handbook of Public Policy* (pp. 449–461). Abingdon, UK: Routledge.
- Carvalho, L. M., & Costa, E. (2014). Seeing education with one's own eyes and through PISA lenses: Considerations of the reception of PISA in European countries. *Discourse: Studies in the Cultural Politics of Education*, 36(5), 638–646.
- Collier, D. (2011). Understanding process tracing. *Political Science & Politics*, 44(4), 823–830.
- Deterding, N. M., & Waters, M. C. (2018). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological Methods & Research*. Published online – doi: 10.1177/0049124118799377
- George, A. L., & Bennet, A. (2005). *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69(S3), S342–S353.
- Jessop, B. (2010). Cultural political economy and critical policy studies. *Critical Policy Studies*, 3(3/4), 336–356.
- John, P. (2003). Is there life after policy streams, advocacy coalitions, and punctuations: Using evolutionary theory to explain policy change? *Policy Studies Journal*, 31(4), 481–498.
- Kingdon, J. W. (1984). *Agendas, Alternatives, and Public Policies*. Boston, MA: Little, Brown.
- Lascoumes, P., & Le Galès, P. (2007). Understanding public policy through its instruments: From the nature of instruments to the sociology of public policy. *Governance*, 20(1), 1–21.
- MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (1998). Codebook development for team-based qualitative analysis. *Cultural Anthropology Methods*, 10(2), 31–36.
- Mason, J. (2002). *Qualitative Researching*. London: Sage.
- Patton, M. Q. (2015). *Qualitative Research and Evaluation Methods: Integrating Theory and Practice* (4th edition). Thousand Oaks, CA: Sage.
- Ritchie, J., Lewis, J., & Elam, G. (2003). Designing and selecting samples. In J. Ritchie & J. Lewis (Eds.), *Qualitative Research Practice. A Guide for Social Science Students and Researchers* (pp. 77–108). London: Sage.
- Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers*. London: Sage.
- Steiner-Khamsi, G., & Waldow, F. (2018). PISA for scandalisation, PISA for projection: The use of interna-

- tional large-scale assessments in education policy making—an introduction. *Globalisation, Societies and Education*, 16(5), 557–565.
- Tansey, O. (2007). Process tracing and elite interviewing: A case for non-probability sampling. *Political Science & Politics*, 40(4), 765–772.
- van der Heijden, J., Kuhlmann, J., Lindquist, E., & Wellstead, A. (2019). Have policy process scholars embraced causal mechanisms? A review of five popular frameworks. *Public Policy & Administration*. Published online – doi: 0.1177/0952076718814894
- 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.
- Waldner, D. (2012). Process tracing and causal mechanisms. In H. Kincaid (Ed.), *The Oxford Handbook of Philosophy of the Social Science* (pp. 65–84). Oxford, : Oxford University Press.

## ANNEX 1: INTERVIEW GUIDE

### Module I: Starting questions

- 1.1. About yourself: could you tell me a little about your trajectory in education, including relevant past positions and the specifics of your current position?

*If not mentioned:* Ask about the academic background.

*If the background is not on education, or is not clear from their previous answer:* Ask how they got involved in education.

- 1.2. Can you describe to me major changes in basic education in the last decade in this country?

[Notice whether they talk about accountability, external evaluations, or use accountability language or rather “quality assurance” language, and adapt the interview to it]

### Module II: Policy formation [one instrument per interviewee, exceptionally two]

Recent reforms in [*name of the county*] have intended to introduce some degree of accountability/quality assurance in education - for teachers, principals and schools. With the following questions, I’m interested in understanding the processes of policy formation that lead to the adoption and development of the accountability/quality assurance system through different policy instruments. We have attempted to summarize the accountability system in this country in the following chart. [Show chart]

- 2.1. Does this chart make sense? Do you think that it represents well the evolution and the current form accountability system in this country? Which of these elements and policies have a bigger impact in ongoing educational processes at the school level? Has any of them that generated more debate than the other? Why?

- 2.2. Could you kindly indicate if you have been involved in the design, adoption or implementation of any of them? If so, in what ways?

*Alternatively:* To what degree, and in what position?) (decision-maker, advisor/expert, consulted party).

*Probe for:* Do you feel you were deliberately excluded from any of them? Why?

- 2.3. Also, which are the elements of the chart you think you are more knowledgeable about? How and why did you get acquainted with each of them? If you were not directly involved, how did you learn about them? Which were your sources of information?

*Probe for:* key informants, secondary documents, hearsay.

### Problem stream

- 2.4. Why do you think this reform/instrument was adopted? Which was the problem intended to be addressed by the adoption of this policy instrument (or educational reform - in case we go for the general reform approach)? Was the problem related with the attitude or quality of the work of teachers?

- 2.5. Can you remember in which sphere and when were these problems discussed for the first time?

- 2.6. What role did research play, if any, in the debate on these educational problems?

- 2.7. Who else contributed to identify these types of educational problems in the country? Was the role of any particular international organization significant?
- 2.8. Would you say there was a general agreement on the causes and relevance of these problems? Or did particular groups, actors or organizations rather hold and raised a different views on that?

### Policy stream

- 2.9. Can you remember when the idea of [*name of the policy instrument or measure*] was discussed for the first time? Do you have any idea about who came up with the idea?

*Alt:* Was this idea spearheaded by any particular group or individual?

- 2.10. [*If the policy was “materialized” in a clear and well-known policy text*] Who worked in the subsequent drafts of the [*decree/reform /bill of education*]?

*Probe for:* locus of policy design – executive power, parliament, appointed committee of experts, international organization, donor organizations, consultants etc.

*Only in government-led processes:* To what extent was the Ministry of Education/Department of Education responsible for the policy design? Were other ministers in the government involved? How so? Are the drafts accessible?

- 2.11. [*Only in government-led processes*] In the process of policy design, which groups or organizations were consulted? Which not? Why?

*Probe for:* inputs from civil society organizations, teachers’ or student unions, etc; ask also how were “relevant” actors selected.

*Probe for:* role of the political parties not in power.

*Probe for:* input from technical staff in the Ministry of Education or other government agencies.

*Probe for:* role of international organizations.

*Probe for:* role of the experts, think tanks, etc.

- 2.12. Can you think of any technical or research documents that played a key role in the process of policy design?

*Probe for:* ad-hoc products; dynamics of instrumentalization, etc.

- 2.13. Can you think of any comparable education experiences (in other regions or countries) that inspired in some way these policy and measures? If so, why do you think these particular areas or countries were considered as valid examples?

*Probe for:* dynamics of policy borrowing, dynamics of instrumentalization and selectivity, influence of particular reference societies.

- 2.14. Did the policy proposals experiment any significant changes during the design process? How do you explain this evolution?

- 2.15. Did any individual or group came up with “alternative” formulations, major amendments?

*Probe for:* inputs from specific actors; concerns or considerations on their feasibility (technical, political, financial).

- 2.16. [Only when the policy instruments have been in place for a significant period of time and the object of criticism] Were any potential problems anticipated at the time of the design of this policy?

### Political stream

- 2.17. Which groups, individuals or organizations were supportive of this policy measure? Do you think that they are shared the same motivations, that is, did they all identify the same advantages? Did they organize in some form of coalition or alliance?

*Probe for:* Had they cooperated or teamed up in the past? Were you part of this coalition?

- 2.18. Which groups, individuals or organizations were opposed to its advancement? Do you think that they are shared the same motivations, that is, did they all identify the same dangers or limitations? Did they organize in some form of coalition or alliance?

*Probe for:* Had they cooperated or teamed up in the past? Were you part of this coalition?

- 2.19. In general terms, do you think that the adoption of this policy/measure was an easy process?

*Alt:* Would you say that was a rather smooth/straight forward, or a rather challenging process?  
[If the process is described as complicated or the interviewee has mentioned actors actively opposed to the policy] What made possible the final adoption of this policy? Can you identify any turning point, which eased the advancement of the reform?

*Probe for:* compromises, political bargaining, communication issues, leadership, triggering event, changes in the government composition.

- 2.20. Which was the degree of consensus within the educational community at the time of adoption?

## Module III. Enactment, autonomy and support

### Awareness of autonomy policies/level of school autonomy

- 3.1. In many education systems, accountability systems is being preceded by higher levels of school autonomy. Do you recognize that this is the case of your country?
- 3.2. How would you define the level of the autonomy of schools in this country?
- 3.3. In which domains do schools have more autonomy? For example personnel, finance, facilities, curriculum, pedagogy, organization of the school, or combinations of these and other domains?
- 3.4. Who is taking the most relevant decisions at the school level? How are decisions on pedagogy and instruction taken, e.g. in full autonomy, after consultation with other bodies, within a restrictive framework set by educational authorities?

### Perception of evolution of school autonomy over time and in recent years

- 3.5. What changes have occurred in the degree of autonomy of schools to take decisions in the last years?
- 3.6. If so, in which domains of authority have these changes occurred more intensively?
- 3.7. How have these changes been received by different actors? Why?

- 3.8. What prompted these changes? Is there a particular policy, reform or program advancing these changes on the autonomy of schools? Have you been involved in the formulation of the school autonomy program/policy [*If so, back to questions on policy formation - Module II*].

### Autonomy to work with the results of the standardized test

- 3.9. What is your opinion on the way test results are communicated to schools, teachers? And to families? In case the school results are published, does this has an impact on parent's school choice decisions, schools status, self-perception, and so on?
- 3.10. And what do you think about the kind of feedback and the nature of the information that these agents receive on the basis of the test?
- 3.11. What are schools expected to do when they receive the results of the [*name of standardized test*]?
- 3.12. Are the expected learning standards sufficiently clear for the schools?
- 3.13. Do you have the feeling that schools react sufficiently to the results? What type of decisions are taken by schools based on [*name of standardized test*]? Who takes these decisions?
- 3.14. Is the data coming from the test useful for educational improvement purposes? What other sources of data would be more useful/are also useful for schools?
- 3.15. What can schools do to improve their results in the test?

### Capacity to work with the standardized test

- 3.16. Is teacher training in your country preparing teachers and principals properly to address the on-going educational and organizational changes/challenges we are talking about?
- Probe for:* Pre-service and in-service.
- 3.17. Do schools have the capacity and the necessary resources to do so? If not, what capacities and resources should be there?
- 3.18. Do schools receive sufficient support to manage the data, and to enact the necessary processes to improve their results? From whom is this support coming? (education department, inspection, private sector).

### Private sector involvement

- 3.19. Are schools and teachers accessing to educational resources (services and materials) coming from the private sector for the purpose of instructional change and results improvement? Which types of resources are being acquired more often?
- Probe for: lesson plans, test preparation simulations, pedagogical advice, teacher training, digital services/technologies.
- 3.20. Do you think that the presence of services and materials for schools coming from the private sector has intensified in the last years? Is so, why do you think it is so?



4.4. Do you consider that the OECD and PISA in particular have an impact in ongoing education debates (SAWA and beyond) in [country name]?

4.5. Which aspects of your country's PISA results (across any survey round), have led to/inspired changes in policy or practice in your country?

*[Need to pay attention to the use of PISA benchmarks, reference to league leaders and factors or characteristics that can explain their success, and mention to negative reference societies].*

4.6. What is your opinion of how education policy formulation occurs in this country? Are the voices of principals and teachers sufficiently taken into account in policy formulation processes?

**Strategies through which policy actors sought to influence the reform process [for non-governmental actors]**

4.7. Do you (or your organization) produce or use any data related to quality in education? If so, do you use this data to inform educational debates in the country?

4.8. Do you use reports, data sources, evidence, etc. coming from other sources to support these strategies? Can you name some of these sources?

4.9. What else does your organization do to influence the education debate in this country (strategies of negotiation, resistance, knowledge mobilization and so on)?

## Module V. Concluding questions

5.1. Is there anything else you think I might be interested in?

5.2. Are there any other people who would suggest I contact about my research?

5.3. If I have any follow-up questions as I review my findings, would you be willing to participate in a very brief follow-up phone call or email exchange?

## ANNEX 2: CODEBOOK

Code	ATTR_POSITION
<i>Brief definition</i>	Current position
<i>Full definition</i>	Current position and associated responsibilities of interviewee.
<i>When not to use</i>	It should not be used when interviewee refers to past positions and/or responsibilities, when interviewee talks about anyone else than him/herself, or when the related events are of no interest for the purposes of the research.
<i>Example</i>	“I am currently working as a Senior Advisor within the Department of Primary and Lower Secondary Education, and my main responsibilities entail ...”.
<i>Group</i>	Attributes of the interviewee

Code	ATTR_BACKGROUND
<i>Brief definition</i>	Professional background
<i>Full definition</i>	Relevant past positions and associated responsibilities of interviewee. It should include any mention to previous job posts, in the current or prior workplaces (also when the interviewee casually drops the fact in the middle of the interview). References to the educational background of interviewee, including the highest level of education obtained (e.g. secondary; university degree; PhD), area of focus of education (e.g. topic of degree), professional development or training (for example during professional career), etc.
<i>When not to use</i>	It should not be used when interviewee refers to current position(s) and/or responsibilities, or when interviewee talks about anyone else than him/herself, or when the related events are of no interest for the purposes of the research.
<i>Example</i>	“Before I started working here, I worked within the Ministry of Education for three years. My main responsibility there was...” “I have a Master’s degree in English (...) and throughout my career I took part in additional training, for example leadership training”.
<i>Group</i>	Attributes of the interviewee

Code	ATTR_AFFILIATION
<i>Brief definition</i>	Political affiliation
<i>Full definition</i>	Political preferences of interviewee or membership to a political party or politically-connoted organizations.
<i>When not to use</i>	This code should only be used when the interviewee talks about his/her own political affiliation or preferences, not when the interviewee talks in more general terms, e.g. “with the recent election, a shift was made to the left”.
<i>Example</i>	“I am a member of the Green Party”; “I do not align myself to the Socio-Democratic political ideology”.
<i>Group</i>	Attributes of the interviewee

Code	SUBJ_SAWA TOC
<b>Brief definition</b>	“Theory of Change” behind SAWA reforms
<b>Full definition</b>	<p><b>Virtues and uses of SAWA policies and instruments:</b> references to the main advantages about that the interviewee identifies with concern to the use of SAWA reform elements. It should include possible education dimensions that would benefit from SAWA tools (equity, quality, efficiency etc.) as well as intrinsic values of SAWA (transparency, meritocracy, fairness, etc.).</p> <p><b>Mechanisms:</b> Causal beliefs of the interviewee about how particular (SAWA) policy instruments, under particular contextual conditions, will trigger particular dynamics that will lead to specific desired effects. In other words, references made by the interviewee to the promises of particular policy instruments (ultimate aims or objectives of the policy instrument), to the mechanisms triggered by SAWA reforms that are to enable the consecution of their ultimate aims, and the contextual conditions that will allow for this chain of events to occur.</p> <p><b>Risks of SAWA policies and instruments:</b> Main disadvantages or risks that the interviewee identifies with concern to the use of SAWA reform elements, in relation to both its inherent logics as well as in relation to its implementation.</p>
<b>When not to use</b>	The code should not be used when the interviewee talks about the working of non-SAWA related policy instruments.
<b>Example</b>	<p>“If one cuts in central regulations around teaching methods, teachers are enabled to adapt their ways of teaching to the individual student, which will lead to greater creativity and dedication. However, this effect may only be expected when the state offers clear framework conditions and provides support and guidance”.</p> <p>“In order to enhance diversification of education, schools need to receive autonomy in pedagogical and organizational terms”.</p>
<b>Group</b>	Subjective perception of SAWA regimes

Code	PROB_CONTENT
<b>Brief definition</b>	Content of the problem
<b>Full definition</b>	<p>Situation requiring attention or questions that need to be solved via a public response and that gain a prominent place on the policy agenda.</p> <p>This includes not only the construction of “brand-new problems” but also the re-interpretation of already recognized issues.</p> <p><b>Possible categories include:</b> Quality problems (related to overall achievement); equity problems (segregation; academic achievement of certain groups); teacher-related problems (poor levels of competence, etc.); administrative culture (excessively centralized or bureaucratized); centralization (need for a more centralized steering); and economic, financial or budgetary difficulties. This is something that could be signaled through a memo.</p> <p><b>NB:</b> the narrative can be used to validate the adoption of pre-established solutions by policy-makers seeking the opportunity for implementation.</p>
<b>When not to use</b>	Use only if the recognition of the problem gain a relatively prominent place on the policy agenda - not when a given issue is perceived as problematic only by the interviewee or related to a “widespread feeling” that never receives public and governmental attention. (This could however be put down in the memo).
<b>Example</b>	“There was growing understanding of the dangers of an under-skilled teacher force. New recruits displayed higher qualifications but their ability when in classroom was poorer than ever as a consequence of a striking lack of practical training”.
<b>Group</b>	Policy process: diagnosis

Code	PROB_DEFINER
<i>Brief definition</i>	Problem-definer
<i>Full definition</i>	<p>References to individual or collective agents responsible for prompting attention to a subject and/or linking a proposal to a problem (thus enhancing their prospects of “moving up in the agenda”).</p> <p>The category is not restricted to those seeking transformative policy change (political coalitions or organizations with an agenda), but also any agent responsible for fixing policy-makers’ attention over a subject (even if they do not push for any particular solution).</p> <p>There might be some overlapping with the Focusing event category – depending on how the interviewee frames the episode (the degree of “agency” he/she identifies in it – e.g., a PISA scandal can be seen both as the result of OECD action but also as a process not animated by anybody in particular).</p>
<i>When not to use</i>	--
<i>Example</i>	“The TALIS results published last year threw light on the limited degree of preparation of our teachers”.
<i>Group</i>	Policy process: diagnosis

Code	PROB_CONSENSUS
<i>Brief definition</i>	Degree of general consensus around the relevance and causes of the problem
<i>Full definition</i>	<p>Comments on the degree of recognition of the problem by others than “original” problem-definers - for instance regarding its centrality within the corresponding policy communities, etc. It should give an idea of the extent to which the perception of the problem is shared by other relevant stakeholders.</p> <p>Can also be used in when the interviewee refers to the degree of acceptance among the general public (published opinion; public feelings).</p>
<i>When not to use</i>	--
<i>Example</i>	“The Conservatives seemed skeptical about the need to grant the federal government additional monitoring responsibilities”.
<i>Group</i>	Policy process: diagnosis

Code	PROB_EVENT
<i>Brief definition</i>	Focusing events
<i>Full definition</i>	<p>Comments on “catalysts” of policy attention: sudden, rare and (perceived as) harmful events that are considered to reveal a policy failure, episodes that open a “window of opportunity” for intensive policy discussion. They may bring less visible problems to the forefront, allow for a reinterpretation of an already recognized issue, or give additional prominence to an already existing problem. The group may include the publication of indicators, catastrophes, etc.</p>
<i>When not to use</i>	--
<i>Example</i>	“The publication of PISA results was a wake-up call for everybody”.
<i>Group</i>	Policy process: diagnosis

Code	POL_DESIGNERS
<i>Brief definition</i>	Locus of policy design
<i>Full definition</i>	References to the set of actors involved in the formulation of policy specifics during the design process. It should include references to the team in charge of developing the policy instrument in the more technical sense – i.e. those with a responsibility in “drafting” it.
<i>When not to use</i>	The code should be used only to identify agents with an influence on specific policy tools (on-the-ground measures rather than general goals pursued by that, or the problems to be solved). It should not be confused with references to key actors in a more general sense (relevant individuals or organizations who pushed for – or conditioned significantly- the policy change).
<i>Example</i>	“The ministry of education formed a consultative commission that came up with a first proposal of the new scholarships program”.
<i>Group</i>	Policy process: policy formation.

Code	POL_EVOLUTION
<i>Brief definition</i>	Evolution of policy proposals
<i>Full definition</i>	Comments on any change undergone by initial policy proposals; changes introduced in each one of the different rounds of policy formulation (substantial changes made to initial proposals and emergence of alternative proposals). This applies to policy proposals still in the process of being developed as well as to fully designed, adopted and implemented programs.
<i>When not to use</i>	The code should not be used to describe the negotiation process but rather to capture these excerpts in which changed made to the policy proposal are tracked. In some cases, this is likely to overlap with comments on negotiation dynamics (in this case, both the POL_EVOLUTION and the POL_NEGOTIATION code should be used). However, in other instances interviewees are likely to offer only “descriptive” insights on how specific aspects of the policy instruments were modified – without mentioning the who/how/why (in this case, only the POL_EVOLUTION code should be used).
<i>Example</i>	“After the parliamentary commission, a new requirement was introduced, so that only non-profit private schools were eligible for receiving public funds. This was very much the result of certain parental associations that pushed for the elimination of the subsidy to any kind of private providers”.
<i>Group</i>	Policy process: policy formation

Code	POL_NEGOTIATION
<i>Brief definition</i>	Negotiation dynamics
<i>Full definition</i>	Mentions to political dynamics contributing to consensus-building and to the advance of a particular proposal, including political bargaining, compromises, concessions, etc.).
<i>When not to use</i>	Does not refer to “persuasion” dynamics in which a given actor genuinely changes their preferences or priorities.
<i>Example</i>	“After two days of massive strike the government withdrew the provisions on early retirement”.
<i>Group</i>	Policy process: policy formation

Code	POL_CONSENSUS
<i>Brief definition</i>	Comments reflective of the consensus around new instruments and policy changes about other actors' rationales and arguments regarding SAWA
<i>Full definition</i>	<p>Degree of consensus about specific policy proposals and divisiveness brought about by the reform within a given policy community, as well as comments on the general public's receptivity (references to national mood, public sentiments/opinion).</p> <p>Should include references about the preferences of other relevant actors regarding SAWA reform (support, opposition, partial support...), as well as the rationales, framing, arguments, and legitimizing narratives used by them during the policy debate.</p> <p>Should include criticisms made by antagonistic forces (unless they were considered as administrative/technical/ideological considerations during the policy design-process).</p>
<i>When not to use</i>	--
<i>Example</i>	<p>"So taxpayers were not keen on seeing their benefits reduced in exchange of nothing... This never gained any sort of public acceptance".</p> <p>"Teachers unions expressed different positions according to their ideological preferences. Unions A and B were definitely in a critical position meanwhile Union C hold a more balanced position".</p> <p>"Initially the conservative party was very reluctant to the reform but since the Ministry was open to their contributions they ended with a more supportive position".</p>
<i>Group</i>	Policy process: policy formation

Code	POL_INCL
<i>Brief definition</i>	Perceptions on in inclusiveness in the policy design process
<i>Full definition</i>	Comments on the openness of the processes to the concerned community, as perceived by the interviewee. Should also include comments on the (perceived) exclusion of certain groups or actors.
<i>When not to use</i>	Should be used only when such an insight is made by the very interviewee (not when the coder reaches a conclusion on that regard).
<i>Example</i>	"There are formal procedures that allow the labor unions to comment on drafts of a policy proposal".
<i>Group</i>	Policy process: policy formation

Code	POL_TIMING
<i>Brief definition</i>	Perceptions on the timing of reforms /back-and-forth dynamics
<i>Full definition</i>	Comments on the "pace" of the policy change (including the identification of drastic vs. incremental processes) and on its reversibility as perceived by the interviewee.
<i>When not to use</i>	The code should be used only when such an insight is made by the very interviewee (not when the coder identifies these dynamics).
<i>Example</i>	"In this country demands for reform are inevitably followed by the appetite for stability but that never lasts. So you are likely to pass a reform, devote some years correcting and refining the whole thing, and then a new reform period and so forth".
<i>Group</i>	Policy process: policy formation

Code	POL_RESEARCH INPUTS
<i>Brief definition</i>	Technical and research inputs considered during the policy design process
<i>Full definition</i>	References to policy briefs, working papers, journal articles, reviews of national policies for education, and another kind of reports and evidence (such data from ILSAs) that has been mobilized to affect the policy design.
<i>When not to use</i>	This code should not be used when the interviewee refers to the evidence that may have contributed to a scandalization process (diagnosis).
<i>Example</i>	“During the policy design process, the World Bank elaborated a very influential report about how to build a Quality Assurance System in Chile”.
<i>Group</i>	Policy process: policy formation; Knowledge mobilization

Code	POL_ECONOMIC CONSID
<i>Brief definition</i>	Economic considerations that make feasible and workable the enactment of SAWA reforms in relation to budgetary capacities and possible financial constraints
<i>Full definition</i>	References to the impact of “material” factors that influenced decision-making processes, especially if policy-makers can adopt SAWA policies because they perceive that such policies are more cost-effective or less expensive than other policy options. Also, it can be used when policy-makers refer to the influence played by the economic environment, the existence of an economic recession, or budget constraints, in the selection of certain policies.
<i>When not to use</i>	This code should not be used when the interviewee refers to financial crises as a triggering factor of an educational crisis (diagnosis).
<i>Example</i>	“The Financial Minister of Chile was in favor of adopting accountability policies because he perceived them as cost-effective measures that are aligned with economic austerity and budget cuts”.
<i>Group</i>	Policy process: policy formation

Code	POL_TECH_ADM CONSID
<i>Brief definition</i>	Technical and administrative considerations
<i>Full definition</i>	<p>This code refers to comments about the administrative and/or technical viability of certain policy ideas, and how these factors influence the policy debate as well as the policy-design process.</p> <p>It should include comments in relation to the design of the test, the production of indicators (validity, representativeness, etc.), the conduction of the test, and the implementation of SAWA arrangements (distribution of responsibilities between concerned authorities; human and material resources required by the policy change; need for new or improved standards and guidelines, etc.).</p> <p>It should also include comments about consideration and anticipation of possible side or perverse effects of SAWA reforms effectively taken into account in the process of policy design.</p>
<i>When not to use</i>	In relation to side effects: it should not be used when possible side-effects resulting from other policies (not SAWA) are mentioned; or when the interviewee expresses concerns of his/her own if they were not taken into account during the policy-design process.
<i>Example</i>	“Unfortunately, we are still a developing country, so such kinds of policy ideas are not technically workable due to the lack of administrative capacity to implement a national large-scale assessment to measure students’ performance”.
<i>Group</i>	Policy process: policy formation

Code	POL_POLITICAL_IDEO_CONSID
<i>Brief definition</i>	Political and ideological considerations
<i>Full definition</i>	References to political and ideological factors and interest taken into account during the design of SAWA reforms, and references about how these factors have influenced the policy debate. It should include mentions to strategic action aimed at affecting how the promoters of (or opponents to) a reform are perceived by the public; or directed at preserving or altering the balance of forces and power. Add references to: ideological, value-related and cultural considerations.
<i>When not to use</i>	--
<i>Example</i>	“The adoption of SAWA reforms allowed us to legitimize ourselves (the Neoliberal Government) in front of the citizens. In their saw bureaucracy as one of the main problems of the education system. So, we promise them to fight against bureau- cracy through the enactment of a high-stakes accountability system with school autonomy. In the end, we won the elections”. “Accountability measures allow both coalitions (center-left and the right) to align their policy preferences, building a great consensus among them. However, they also have some latent differences: for example, the center-left coalition was in favor of introducing a Superintendency to take control over the management of school’s financial resources, while the right-wing coalition was against the introduc- tion of such measure within the Quality Assurance System”.
<i>Group</i>	Policy process: policy formation

Code	ACT_CENTRALITY
<i>Brief definition</i>	Key and influential agents
<i>Full definition</i>	References to stakeholders perceived and identified by the interviewee as the key or influential agents that participated in the policy process (government, oppo- sition, think tanks, policy entrepreneurs, political leaders, mass media, teachers’ union, students’ movement, social movements, and so on) or as having displayed some degree of leadership or inspirational role in the promotion and the advance of SAWA reforms (as well as over the balance of forces between different coalitions). It may include (a) core actors in the governmental sector (executive staff, central agencies and task forces); (b) public sector insiders (commissions, research council- is, international organizations); (c) private sector insiders (“formally” invited by po- licy-makers) (consultants, political party staff) and (d) non-governmental outsiders (business groups, trade unions, academics, think tanks, media, etc.). It might also include references to actors’ sources of influence - references to the sources of influence of the interviewee that would have enabled them to effecti- vely determine or change the course of policy change (including their prestige and reputation, financial resources, social capital and connections to other agents, as well as external circumstances that created an environment receptive to the action or ideas of these agents). (This might be difficult to code since it is frequently not made explicit by the very interviewee).
<i>When not to use</i>	The code should not be confused with the locus of the policy (policy designers).
<i>Example</i>	“The alliance between the right-wing Government and the Christian-Democratic party allowed the enactment of the SAWA reforms in Chile”. “They were the most influential actor in the design of the education reform act”.
<i>Group</i>	Policy process: balance of forces, actors

Code	ACT_COOPERATION
<i>Brief definition</i>	Cooperation between actors
<i>Full definition</i>	References to sporadic or reiterative modes of collaboration between different actors in order to advance education reform according to some degree of shared or compatible policy (or/and political) preferences.
<i>When not to use</i>	--
<i>Example</i>	“Students and Teachers Unions shared a common strategy of opposition based on social mobilization”.
<i>Group</i>	Policy process: balance of forces, actors

Code	ACT_STRATEGIC ACTION
<i>Brief definition</i>	Strategic action
<i>Full definition</i>	References to the repertoires of action deployed by key agents in order to advance their policy goals and preferences. This includes a range of public and non-public activities, such as (but not limited to): knowledge mobilization strategies (commissioning, conducting, disseminating research; media releases, public awareness campaigns; etc.); lobbying; purposive inter or intra-organizational change (appointing a person explicitly focused on the issue, joining or creating an issue-specific coalition, etc.); formal and informal lobbying; street and industrial action (demonstrations, strikes, etc.).  The codes include strategic actions oriented to condition and influence policy in any sense (ensuring its progress; modifying particular provisions, etc.) and also apply to the opposition and resistance to policies or policy tools.
<i>When not to use</i>	The code should not be used to describe strategies whose “materialization” is unclear.
<i>Example</i>	“We published a white paper which was very influential in defining the priorities of education reform”.
<i>Group</i>	Policy process: balance of forces, actors

Code	GEN_REFERENCE SOCIETIES
<i>Brief definition</i>	Reference and counter-reference societies and/or sectors
<i>Full definition</i>	Comments on so-called “reference societies” (and counter-references societies) at any point during the policy-shaping process (including diagnostic action as well as policy-design practices).  Reference societies are those countries (e.g., ILSAs league leaders) likely to appear in scandalization processes or to be used to showcase desirable policy solutions that should be emulated and enacted. Counter-reference societies are usually mentioned with some heterostereotypes to avoid the adoption of certain reforms.
<i>When not to use</i>	It should not be used to code quotes about the influence of ILSAs.
<i>Example</i>	“The design of our accountability system was inspired by countries such as England, New Zealand and the US”.  “The example of Sweden shows us that the combination of school choice and a voucher system can have damaging effects in terms of education quality and equity”.
<i>Group</i>	Policy process: policy formation; knowledge mobilization

Code	GEN_INSPIRATION
<i>Brief definition</i>	Nature of the “emulation process” (or of a process presented as an “emulation”); uses of reference societies or inspirational models
<i>Full definition</i>	References to the specific ways in which a policy model was adopted. With this code we aim at capturing if the “imitation” dynamics affect the practical implementation of the reform (policy-design) or are rather used to frame and guide the spirit of the policy idea (i.e., related to the to the adoption of a new paradigm). The code could also be used in relation to other policy sectors deliberately used as an “example” or inspiration for education reforms.  <b>NB:</b> This code would in fact include references societies but we found it useful to codes references societies as a separate code.
<i>When not to use</i>	--
<i>Example</i>	“This reform is inspired by new public management paradigm”.
<i>Group</i>	Policy process: policy formation; knowledge mobilization

Code	GEN_FUTURE
<i>Brief definition</i>	Expectations regarding next or future developments of the policy
<i>Full definition</i>	References to conjectures, predictions and expectations regarding future developments of the policy instrument or arrangements; comments on the probable evolution of the policy discussed.
<i>When not to use</i>	The code should not be used when the interviewee comments on what he/she perceives as desirable future developments, but only when he/she discusses a <i>likely</i> turn of the events.
<i>Example</i>	“I guess the Minister will eventually bend to at least some of the union’s demands, so they probably will be giving greater weight to the peer evaluation component”.
<i>Group</i>	General; Other

Code	GEN_ENACTMENT
<i>Brief definition</i>	General comments on the enactment of the policy measure
<i>Full definition</i>	References to the actual enactment of the policy measures discussed in the interview – how they have been put in practice, interpreted or renegotiated by policy “enactors” and how this has impacted the eventual policy outcomes. Special emphasis should be put on comments regarding the possible “gap” or disjointedness between the original expectations and theory of change and the actual effects of the policy “in context”.
<i>When not to use</i>	--
<i>Example</i>	“So the inspectorates became responsible for the supervision of the new scheme... but as it turns out most of them do not have so much knowledge about what was going on with in-service training, so they basically rely on principals’ self-reporting, and that makes the whole thing pretty circular, in fact”.
<i>Group</i>	General; Other

Code	GEN_POL EFFECT
<i>Brief definition</i>	General comments on the effects of the policy measure
<i>Full definition</i>	References to any comment on the (perceived) consequences of a given policy or policy instrument - including changes in any education dimension (quality, equity, effectiveness, etc); as well as unintended or side effects.
<i>When not to use</i>	--
<i>Example</i>	“Student achievement improved hugely as a consequence of the new policy”.
<i>Group</i>	General; Other

Code	GEN_POLICY INSTRUMENT
<i>Brief definition</i>	Policy instrument, measure or program discussed by the interviewee
<i>Full definition</i>	References to the policy tool(s) that center the interview and constitute its main focus.
<i>When not to use</i>	--
<i>Example</i>	--
<i>Group</i>	General; Other