

GABRIEL CHOUHY

*The moral life of econometric equations:
Factoring class inequality
into school quality valuations in Chile*

Abstract

Recent sociological scholarship on market design is ill-equipped to understand the normative and political aspects of experts' practices in connection to political conflicts over the commodification of social rights. I develop an original approach to the politicized use of market devices to address collective concerns in a noneconomic policy field: education. When designing a high-stakes school accountability system, policymakers in Chile confronted a moral conundrum: should schools be valued according to their students' absolute proficiency, or according to the school's relative effectiveness? Progressive and conservative experts in charge of settling this dilemma pushed for using the statistical model (OLS vs. HLM) that yielded rankings that fit their moral preferences. Through qualitative analyses of experts' real-world application of quantitative methods, as well as experts' interpretations of these methods' performative consequences, I mobilize the much-debated concept of "moral background" to unravel the conditions for subsuming ideological dissent into consensual forms of decision-making.

Keywords: Accountability; Market devices; Experts; Morality; Neoliberalism.

Introduction

SINCE NEOLIBERALISM took root, market mechanisms have encroached on the organization of policy fields for the provision of public

This research was supported by the National Academy of Education/Spencer Foundation, the Mellon Foundation, and Center for Latin American Studies at the University of Pittsburgh.

I especially thank Lisa D. Brush, John Markoff, and Samantha Plummer for their

generous revisions and suggestions. Aníbal Pérez-Liñán, Sean Kelly, Sean Reardon, Gabriel Abend, Maria Akchurin, María José Álvarez, José Ossandón, Kieran Healy, and four anonymous reviewers also provided insightful comments.

141

Gabriel CHOUHY, Tulane University [gchouhyalgorta@tulane.edu].

European Journal of Sociology, 62, 1 (2021), pp. 141–182—0003-9756/21/0000-9000\$07.50 per art + \$0.10 per page
© *European Journal of Sociology* 2021. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence [http://creativecommons.org/licenses/by/4.0/], which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.
doi: 10.1017/S0003975621000059

goods such as transportation, healthcare, education, or social security. But oftentimes marketization fails to deliver satisfactory outcomes, inviting experts and regulators to assess market failures and rewrite their organizing rules. An emerging scholarship within the social studies of markets tradition is now theorizing about “the organization of markets for collective concerns and their failure” [see e.g., the special issue of *Economy and Society* edited by Frankel, Ossandón and Pallesen 2019]. Instead of focusing on political junctures in which market-minded policymakers invoke the general principles of privatization, choice, and competition to oppose central planning within malfunctioning public bureaucracies, these authors look at processes whereby already marketized policy fields become problematized, evaluated, and fixed to correct socially undesirable market failures. A remarkable conclusion drawn from multiple case studies is that little room remains for a politics that includes “the possibility of problematizing existing policies in ways other than as poorly functioning markets” [Frankel, Ossandón and Pallesen 2019, 166]. Market-minded policies and devices have not become immune to evaluation and critique, but, after neoliberalism, the type of knowledge generally mobilized to problematize market-enhancing policy instruments is now much more circumscribed to the actual functioning of these devices and much less inimical to the market organization of policy fields themselves. Neoliberal doxa [Mudge 2011, 2008; Amable 2011] now dictates technocratic common sense.

This paper contributes to the vital question posed by Ossandón and Ureta [2019, 176] with respect to neoliberal resilience, namely, “How does a critical evaluation of market based policies, rather than triggering a movement, for instance, toward fundamentally different modes of organizing a given area, end up consolidating markets as policy instruments?” Like these authors, my empirical focus is on experts’ work in evaluating and repairing market devices that failed to deliver optimal policy outcomes. However, I give especial attention to the way in which external demands for decommodification shape this technical work. Bringing politics back into the social studies of markets, I argue that, contrary to the idea that neoliberal doxa is fully entrenched, the technical process of market evaluation and repair remains enmeshed in ideological conflict over the moral virtue of marketization. To do so, I lay out a reflexive approach to the politicized uses of social scientific expertise (especially, but not only, economic expertise) for regulatory decision-making within education, a policy domain the sociological literature has largely overlooked despite being increasingly subject to top-down rationalization by technocratic and market forces worldwide.

My case study focuses on the institutionalization of test-based, high-stakes accountability in Chilean education and, specifically, the design and implementation of the so-called School Ordinalization Methodology (SOM)¹—an official categorization of school quality set out to perform socially optimal market outcomes by *fairly* holding providers (schools) responsible to consumers (students). In response to mass student protests against the educational inequities produced by the so-called “market model” of education, the Chilean Congress began instituting new regulations aimed at holding both municipal and voucher-funded private schools accountable to quality standards. Thus, the Education Quality Assurance Law of 2011 transformed education governance into a system of specialized agencies in charge of setting, measuring, overseeing, and enforcing these standards. Importantly, lawmakers also mandated the “fair responsabilization” of schools to address mounting criticism that school quality metrics were oblivious to the severe socioeconomic segregation of the school system, and therefore stigmatized public schools with high concentrations of low-socioeconomic status (SES) students, fostering exit to the voucher-funded private sector². Yet, ambiguity in what precisely “fair responsabilization” meant confronted the education experts in charge of devising an objective, socially sensitive ordinal metric of school quality with a moral conundrum, one that pervades the use of benchmarking rules for high-stake accountability purposes: on the one hand, is it fair (to schools) to assess school effectiveness on the basis of factors for which the schools are not responsible, e.g., SES? On the other hand, is it fair (to students) to set multiple proficiency benchmarks, that is, different standards based on their socioeconomic characteristics?

I take the “valuation problem”—the definition of criteria concerning “the assignment of values to heterogeneous products within the same market” [Beckert 2009, 254]—education experts faced when going about settling the fair responsabilization controversy as an opportunity to integrate and expand existing sociological knowledge on the politics of markets for collective concerns. Specifically, I seek to make sense of a twofold paradox emerging from this controversy that the existing literature is ill-equipped to understand: on the one hand, why, despite

¹ The Spanish word “ordenación” relates to the verb “ordenar,” which, like the English verb “to order,” means both to give instructions or requests, and to arrange objects or elements into groups. To avoid confusion over the cognate “ordination,” with its misleading connotation of taking or conferring holy

orders, I chose to translate “ordenación” into “ordinalization,” following Fourcade’s [2016] theorizing about the “sociotechnical channels through which ordinal judgments are now elaborated” [175]. The SOM nicely fits this definition.

² HIRSCHMAN 1970.

appeals to objectivity and impartiality, did the experts weaponize their calculations to suit their subjective conceptions of market education; on the other, how, notwithstanding politicization, did they still manage to ground their disagreements in shared technical procedures and successfully implement the SOM, transforming education quality into an objective, quantifiable, and politically consequential feature of market governability?

To address this twofold paradox, I draw upon the sociology of markets and morality, conceptualizing the SOM as a technology of economic valuation that, as such, embodies an intensely moralized technopolitical project. The quest for fairness in measuring school quality in Chile speaks directly to scholarship that examines the morally controversial valuation of goods such as health, nature, love, or human life. Education fits perfectly into the category of “enrichment” [Boltanski and Esquerre 2016] or “peculiar” [Fourcade 2011] goods, the valuation of which typically implies moral quandaries. That the economic status of education as a market good may itself be widely contested means that its commensuration immediately yields claims on incommensurables [Espeland and Stevens 1998], and invites public invocations of irreducible orders of worth [Boltanski and Thévenot 2006]. The appraisal of so unique an object as school quality may easily overlap with aesthetic judgements [Karpik 2010], or simply instantiate preexisting status hierarchies of both producers and consumers [Podolny 1993].

I therefore see the technical construction of valuation devices as embedded in a broader conflict over the moral virtue or evil of markets (in this case, education markets). But this conflict, I show, concerns not only the value of the goods at stake (e.g., private vs. public education), but also the value of the valuation devices themselves, i.e., how much these devices should serve the political purpose of organizing the provision of these goods in a market-like fashion. The fair responsabilization controversy led to a clash between conservative and progressive experts. Though the conservative government staffed the commission that drafted the first ordinalization proposal with market-friendly experts, and took up the majority of seats in the new regulatory agency tasked with implementing the SOM, an ideological battle over the policy instrument’s inherent virtue as a market-enhancing device nonetheless transpired. Skillfully handling this disagreement, the experts reduced it to a technical decision between econometric models (OLS vs. HLM) that empirically identify school quality as a statistical parameter. Experts’ conflicting moral judgments on the intrinsic value attributed to the SOM as a tool for consumer choice and market competition nevertheless

grounded their adjudication between regression equations. While conservatives advocated for the model (OLS) that would make the SOM more intuitive for parental choice and better at “nudging” schools (whether private or public), progressives showed little regard for the behavioral properties of the SOM and instead pushed for the model (HLM) that would better compensate public schools for the “quality handicap” associated with educating larger numbers of disadvantaged students.

Still, despite this moral divide, my analysis shows that the experts eventually succeeded in transforming the chosen metric into an impartial, presumably fair, and politically consequential tool for ordinalizing the quality of all Chilean schools. To account for this second half of the paradox, the paper elaborates on Abend’s [2014] concept of “moral background”, shedding light on the practical uses of social-scientific expertise in the reorganization of markets for collective concerns. Instead of asking why first-order agreements can be possible even under very different moral backgrounds, however, I turn Abend’s framework upside down, seeking to establish how second-order moral conventions may facilitate political consensus even in contexts of overt ideological dissensus. In this sense, I treat experts’ agreements and disagreements over the “fairness controversy” as nested in distinct orders of morality that are relatively independent and yet inextricably connected. While disagreements over the role of market devices in the organization of the education policy field belong to a first-order morality, I contend that agreements of a second-order have more consequential implications. Chilean technocrats shared styles of reasoning—i.e., “collections of orienting concepts, ways of thinking about problems, causal assumptions and approaches to methodology” [Hirschman and Berman 2014, 794, see also Hacking 2002]—that enabled the commensuration and adjudication of a priori irreducible moral choices. Experts’ common subscription to scientific and technical conventions therefore provided the second-order cognitive tools necessary to effectively settle moral disagreements of first order. Constitutive of technocratic forms of policymaking, this “econometric” moral background proves a pivotal mechanism of neoliberal resilience.

I proceed as follows. The next two sections briefly situate the SOM in the international context, summarize its characteristics, and justify it as a case worth analyzing through the sociology of markets perspective. A methodological note follows. The fifth section contains the bulk of my empirical analysis: a brief historical account of the SOM’s origin, followed by an in-depth examination of how regulators in charge of the new accountability system handled the moral dilemma that arose when

operationalizing the legal mandate that school quality scores not be confounded with students' SES. I conclude by discussing my main theoretical contributions in light of my findings.

The School Ordinalization Methodology in Context

The steady increase in standardized testing for the external evaluation of teachers and schools is perhaps the most salient international trend in education policy of the last four decades. Standards-based evaluation policies in education were first institutionalized in the 1980s in liberal-market economies such as the US and the UK, but have henceforth spread across education systems worldwide as part of a global shift towards quasi-market forms of social service delivery and the diffusion of new public management paradigms emphasizing autonomy, competition, and outcomes-based (instead of process-based) accountability. Institutional reforms have underpinned this process, with governments embarking on the creation of specialized regulatory agencies tasked with developing both evaluation expertise and enforcement capacities [OECD 2016, 2013].³

The quest for high-stakes evaluation regimes, scholars have argued, enables technocratic logic to penetrate the field of education, rationalizing schools from the top-down as to make them governable at a distance [Mehta 2013; Gorur 2016; Grek 2009]. Two intertwined and mutually reinforcing developments critically contributed to this overall trend. First, the consolidation of a specialized field of education knowledge—with education economics taking a leading role—transformed the conceptualization of human capital from an *input* of economic growth or earnings functions (e.g., skilled labor) to an *output* of education production functions (e.g., academic achievement) in which both teacher and school quality can be factored as measurable inputs. Second, the growing recourse to economic experts and devices by international organizations, public bureaucracies, and regulatory agencies produced a paradigmatic shift from equity to efficiency in education policymaking [Griffen 2020; Teixeira 2000; Mangez and Hilgers 2012].

³ Though national evaluation regimes have become ubiquitous, not all countries use testing for high-stakes purposes. A recent survey of twenty OECD countries reveals that while testing increased across the board between the 2000 and 2015, only eleven countries make extensive use of testing for school

accountability. Likewise, testing is consequential for school closures in nine of these countries, and only in three are test scores widely used for external comparison via rankings or league tables [BERGBAUER, HANUSHEK and WÖESSMANN 2018].

Increasing emphasis on outcomes and efficiency in turn has brought to the fore demands for equity in test-based assessments of school performance. Testing organizations behind popularized metrics are now reconsidering their methods to accommodate mounting demands for fairness. To encourage context-sensitive assessments of student preparedness by selective college admission offices, the College Board, a flagship educational testing company in the US, recently began reporting socioeconomic information on students' neighborhoods and high schools along with their test results. Likewise, to better reflect schools' efforts at improving the learning outcomes of underserved students, the organization that produces the neighborhood schools' quality metric used by the most popular real state online portals in the US has recently boosted the weight of growth and equity measures, in light of evidence that the unadjusted scores used in housing transactions increased segregation [Hasan and Kumar 2019]. This quest for metrological fairness is now pervasive, cutting across institutional landscapes and national assessment cultures. And a commitment to equity is almost unanimous among experts, to the point that even enthusiastic academic exponents of high-stakes testing recommend the use of adjusted measures that are sensitive to both prior achievement and other contextual factors [Hanushek 2009; Harris 2011], though the statistical complexity of these adjustments reduces model transparency and thereby invites further equity concerns.

In Chile, the institutionalization of high-stakes testing occasioned the adoption of standards-based accountability rules and the subsequent search for a way to equitably mete out the punishments and rewards these rules prescribed. Under the mandates provided by the Law of Education Quality Assurance (LEQA), in March 2014, the Chilean government issued an executive order enacting the "Ordinalization Methodology of all Educational Establishments Recognized by the State."⁴ The decree details all the indicators, data sources, decision rules, equations, and estimation methods the Agency of Education Quality (AEQ) shall use to classify schools into four "performance categories" (high, medium, medium-low, and insufficient), obtained by partitioning an interval scale of education quality. Furthermore, the regulation stipulates that the ranking position of a particular school be computed by subtraction: the effect (aggregated at the school level) of a set of student-level variables of SES and other characteristics of the school context,

⁴ "Metodología de Ordenación de Todos los Establecimientos Educativos Reconocidos por el Estado." See [http://www.leychile.](http://www.leychile.cl/Navegar?idNorma=1060182)

[cl/Navegar?idNorma=1060182](http://www.leychile.cl/Navegar?idNorma=1060182), retrieved on January 19th, 2017.

minus the predicted student test-scores (also aggregated at the school level) (see equation 16 in the decree). In turn, the two terms of this subtraction—a routine post-estimation adjustment—can be obtained through an ordinary least squares (OLS) regression at the student level that predicts the “unadjusted index” as a function of a set of socioeconomic variables, plus an error term that captures unobserved heterogeneity in student performance (see equation 15).

Strikingly, the decree refers to this error term not to assume its normal distribution, homoscedasticity, exogeneity, and non-autocorrelation—that is, the statistical features that help analysts gauge the likelihood that the estimated coefficients and standard errors in the model are relatively unbiased in the technical sense. Rather, the decree *rules*—and thus states performatively—that this error term expresses the *true value* of schools.

In equation N° 15 the individual student score consists of two parts. The first is an estimate of the student score that this methodology attributes to her characteristics and context $c + \sum \beta_{CA} \times CA_{NI}$, which corresponds to the average score of those students that share the same level of student [socioeconomic] characteristics at the individual level. The other part of equation N° 15, ε_{NI} , corresponds to the fraction of the student score that is above or below that of an average student with the same level of student [socioeconomic] characteristics at the individual level. *It will be considered that this fraction of the student's score is not attributable to her context, so it will be considered of the school's responsibility.* [author's translation and italics]

This regulation epitomizes the rationale reigning within the governance bodies in charge of enforcing high-stakes accountability policy. The decree assumes a sharp distinction between providers and external regulators. Schools, all publicly-funded but run either by municipalities or private organizations, provide education services. Rating agencies assess the “value” of the work these schools do, keep families informed about differences in service quality, and allocate positive or negative sanctions and/or support to the schools they rank. Confident enough in their capacity to represent differences in service quality through a single metric and seeing the latter as a function of a transparent combination of independent measurable inputs—“the education production function”, to use the language of education economics [Todd and Wolpin 2003; Hanushek 1986]—the regulators declare by fiat that schools be liable for the variance in test-scores not accounted for in the “class effects” part of the model. Hence the assertion that the error term of the estimated OLS regression is but a non-observable measure of school effects on achievement.⁵

⁵ Statistically controlling for student and contextual factors does not suffice to produce unbiased residual estimates of school effects, unless two conditions are met: a) the

estimation method considers the hierarchical structure of the data (students nested in schools); b) students are randomly allocated to schools. In the absence of pure experimental

Moreover, what the LEQA stipulates the state must observe and manipulate—statistically, albeit not in practice—is the multiplicity of social conditionings that class inequality exerts on student performance:

[The ordinalization] of schools will consider the student’s characteristics, including, among others, her *vulnerability*, and, when appropriate, *indicators of progress or value added*. All in all, gradually, the ordinalization of schools will tend to be performed *independently of the socioeconomic characteristics* of the students, to the extent to that the [quality assurance] system corrects the achievement gaps attributable to such characteristics. [author’s translation and italics]

Policymakers’ most elemental aspiration is to discount SES effects from assessments of school quality, measured as the degree of compliance with *proficiency* standards. In a second stage, the expectation is that adjusted quality assessments also include student growth measures to better approximate statistical estimates of *effectiveness*, understood as the school’s contribution to students’ achievement⁶. In the long term, however, the goal is to progressively deindex school quality valuation from socioeconomic factors, insofar as the incentives set in motion by accountability regulation succeed in reducing the SES achievement gap to a minimum. Meanwhile, the method of holding schools accountable for the service quality they offer must account for “class handicaps.” Not accounting for SES might be unfair; it might unjustly punish schools and families in “vulnerable contexts.” And, yet, the paradox is that the state will nonetheless “value” schools for whatever they do that remains, precisely, unobserved by the valuation device the state is putting in place to bring order and stability to the school market.

Towards a Sociology of School Quality Valuation Devices

What motives were behind the implementation of this peculiar device that explicitly discounts SES from measures of school performance and then assigns quality values to all that goes unobserved? As my analysis will make clear, the SOM resembles other cases of post-marketization

conditions, relatively unbiased estimates of school effects can only be obtained if specific school practices are made observable and tested as fixed effects, which requires a theory of how schools work [see RAUDENBUSH and WILLMS 1995]. None of these conditions is met in the case of the SOM.

⁶ Value added models require panel data on students’ test scores. Although the use of these methods in high-stakes accountability systems has expanded significantly in the last decade, both the Statistical Association [2014] and the American Educational Research Association [2015] have warned about their potentially adverse consequences.

policy tools arranged to address a collective concern. Yet, the fact that quality assurance regulation in Chile stemmed from student protests against market education after decades of far-reaching privatization and decentralization makes the enactment of the SOM a somewhat exceptional case as far as education policy is concerned. In the United States, for instance, high-stakes accountability progressed faster than other pro-market policies. While bipartisan consensus made testing and ranking the leitmotif of education reform, advocates of large-scale privatization never achieved a legislative majority [Mehta 2013]. Indeed, since the enactment of No Child Left Behind, accountability has worked as a Trojan horse for pro-market reformers to advance vouchers, charter schools, and choice. Similarly, in England—another early adopter of market-based approaches to education—the use of league tables of school quality became widespread as early as the 1990s. Only recently the government began using test-based performance metrics to promote the conversion of underperforming public schools into privately managed academy trusts. In contrast, policymakers in Chile predicated accountability rules—with the SOM taking centerstage—on the need to fix an already deregulated, largely privatized school market (by the year the LEQA was enacted, private voucher schools accounted for more than half of national student enrollments).

The SOM is therefore a case of what sociologists [Ossandón and Ureta 2019] duly note as a second (post-neoliberalization) moment in the way experts conceive of and practically use market devices. This second moment entails a shift from the standardized description of the market mechanism as an abstract, quasi-natural entity opposed to planning and rational organization, to the technical engineering of problem-solving market devices aimed at delivering specific policy and organizational outcomes. Rather than a panacea for politicized and opaque bureaucracies, markets are now used as organizational tools that “could be skewed in the outcomes favored” [Nik-Khah and Mirowski 2019: 282]. Hence the cleavage that structured 20th century political struggles—and that the neoliberal movement so actively enhanced—has given way to a situation in which the formal distinction between market-like and bureaucratic-like forms of organization is increasingly blurred. Conceptually, the institutionalization of ordinal categories of school quality in Chile can thus be seen as both an attempt at bureaucratic rationalization (holding schools accountable to government standards) as well as an effort at market steering (creating the right incentives to empower parents and make providers compete on the basis of service improvement).

At the backstage of this overall transformation in the political use of markets for organizational advancement, the literature identifies the growing influence of specific kinds of experts (e.g., information, behavioral, or experimental economists), who, laying claims to jurisdiction on noneconomic policy fields, mobilize the tools of market engineering (e.g., accountability metrics, matching algorithms, centralized auctions, behavioral nudges) as optimal technological solutions [Ossandón and Ureta 2019]. Importantly, this represents a form of expertise that naturalizes the organizational form of the market as the only game in town. Everything proceeds as though the experts' decisions were only about different alternatives for market reorganization [Frankel, Ossandón and Pallesen 2019, 166]. In detailing how markets are evaluated and transformed with the specific tools of economics expertise, the social studies of markets scholarship showcases technocracy's decisive contribution to neoliberal resilience. Institutionalized as "a policy regime where market-based policy instruments are actively constructed by governments to act upon collective problems" [Ossandón and Ureta 2019, 176], neoliberalism proves resistant to demands for de-commodification inasmuch as the "market engineers" behind these instruments problematize them in a way that precludes market reversal or even discussion of non-market solutions. Implicit in this theorizing is that neoliberalism's entrenchment hinges on the perseverant vigilance of market-minded experts whose scientific judgement, pace Polanyian-like moral critique of market society, dictates that surgically designed market fixes remain technically superior—hence morally preferable—for delivering socially desired outcomes. Acting as enlightened gatekeepers, technocrats come to embody Davies' [2016: 6] definition of neoliberalism as "the pursuit of the disenchantment of politics by economics."

Focusing on the experts' task of making the market work as a "good" market, social scientists have revealed how technocratic critique of market-based policies paradoxically leads to market reinforcement instead of reversal. Yet, the literature fails to explain why the value judgements supposedly occluded or even erased by marketization—including moral disagreements over the question of the market organization of policy fields—sometimes (as was the case of the SOM) reemerge in a technical fashion, let alone how, despite the politicization of what is not supposed to be political (i.e., the "technical"), the prevailing form of market expertise manages to preserve its performative efficacy (as it did after the SOM experts settled the fair responsabilization controversy). My empirical puzzle, therefore, comprises two apparently contradictory aspects in need of simultaneous elucidation. First, experts' *failure* to

depoliticize demands for fair responsabilization in market-oriented education, which points to the moral and ideological embeddedness of their scientific judgement. Second, experts' *success* in imbuing their market device with the technical capacity to perform presumably objective valuations (as finally did the SOM) even in highly politicized contexts, which points to the conventions and procedures that make expertise efficacious in the first place.

Elaborating on Abend's [2014] concept of "moral background", this paper addresses both of these aspects holistically. I contend that Abend's distinction between nested and yet intertwined orders of morality makes it possible to discern the conditions for subsuming ideological dissent into consensual forms of technocratic decision-making. Scholars have criticized the "moral background" framework for its limited generalizability beyond the two ideal types for which it was originally developed: the "Christian Merchant" and "Standards of Practice" types of business ethics. Others have posited that because the defining features of the moral background sustaining the "Standards of Practice" type (e.g., scientism, naturalism, utilitarianism) have become virtually indisputable in contemporary economic discourse, Abend's guiding thesis—that first-order moral agreements may be nevertheless grounded in irreducible regimes of moral justification—lacks sociological currency. Specifically, the concept has little to say about opposite ways of rationalizing business conduct (e.g., shareholder vs. stakeholder value) that, albeit rooted in a similar moral background, reflect radically different views of how capitalism should be organized.

Taking heed of said critiques [for an overview, see e.g., Livne *et al.* 2018], this research extends Abend's framework to contemporary instances of technocratic decision-making, yet reverses its analytical logic. Instead of asking how incommensurable moral backgrounds may lead to concurring moral stances, I demonstrate how technocrats' shared styles of going about translating moral disputes into technical decisions can make consensus possible even under conditions of moral dissent. I thus draw upon Abend's analytical distinction between first and second orders of morality to theorize the conditions of commensurability of competing moral judgments among the experts in charge of devising the SOM. A first-order morality concerns questions such as what different moral views exist about something and what kinds of behaviors are associated with different moral standpoints. A second-order morality, on the other hand, concerns more fundamental issues such as what type of objects can be morally evaluated, what kinds of reasons can be legitimately invoked to support these evaluations, what repertoires of concepts

and ontologies ground such reasons, or whether moral arguments themselves can be arbitrated by recourse to objective procedures. Together, these second-order issues take the form of a shared set of tacit conventions, constituting a common moral background.

As the empirical section makes clear, when devising the SOM, conservative and progressive technocrats clashed over the selection of the statistical method—ordinary least squares vs. hierarchical linear models—yielding ordinalizations that better fit their partisan definitions of fairness. A close reading of experts’ reason-giving shows that what divided their modelling choices was the intrinsic value they attributed to the SOM as a tool for consumer choice and market competition. However sharp, experts’ discrepancy in the fairness controversy was but a first-order disagreement, one that reflected morally charged evaluations of the effectiveness of market mechanisms in the provision of public goods. Still, a common second-order morality facilitated the conditions for the market device’s performative efficacy. Beyond the surreptitious ideological dispute among experts, the divide was not deep enough to splinter a more fundamental consensus, one which allowed them to settle the controversy and get on with creating the SOM. Although their moral judgments on the distributive consequences associated with alternative ordinalizations diverged, they ultimately shared an economic style of reasoning [Hirschman and Berman 2014]: an array of tacit conventions—econometric conventions in particular—that aided in reducing the moral conundrum to a tradeoff between proficiency and effectiveness standards. It was the epistemological assumptions, routine measurements, administrative data, and calculation methods assembled in the process that instituted education quality as an objective, quantifiable trait present in all schools in different degrees and, crucially, independent of social context.

Research Methods and Data

To open the “black box” of the SOM as a valuation device, I drew on Latour’s [1987] perspective for studying “science in action.” My method, however, is not ethnographic. Though my goal was to analyze “the SOM in the making”, in practice, I did not “follow education specialists around”—and “back into their research lab.” Still, my inquiry did not simply focus on the SOM’s enactment as a “ready-made” market device. Neither did it take post-hoc justifications of the SOM as evidence

of expert consensus. Rather, I considered the executive order enacting a particular ordinalization method as well as the official justifications that supported it as culminating statements in a sequence of expert performances. Precisely because these statements closed—and black-boxed—the fair responsabilization controversy (or at least settled most of its technical aspects), I looked back into the sequence to reconstruct the process through which an issue, i.e. market failure in the provision of quality education, was first made into a problem, then transformed into a quest for measuring school quality fairly, and finally reduced to a technical debate on how this fairness requirement could be satisfied through different modeling techniques. In this reconstruction, another important statement stood out: the public record of the session where the AEQ board members voted and argued for the statistical model used to factor the student socioeconomic status into estimates of school quality. Exposing the Janus-faced character of the ready-made SOM vis-à-vis the SOM in the making, this piece of evidence allowed for politics to reemerge in full force.

Hence the bulk of my analysis consisted of a post-hoc, second-order problematization of experts' real-world application of quantitative methods, as well as experts' interpretations of these methods' performative consequences. I paid particular attention to when, how, and why these experts resorted to these methods to settle the fair responsabilization controversy, examining the decision-making process through which they deemed some facts and criteria about school assessment useful or useless and how they rendered effective some valuation criteria and not others. I also looked at how the relevant people with an authoritative say in matters of school assessment were enlisted and convinced. In short, I sought to understand how, when all these people, facts, and methods hold together, the quality assessments performed by the SOM became accepted as objective reality.

To further contextualize this analysis, I delved into an archive of legislative acts, policy documents, official minutes, public statements, and reports I collected and sorted chronologically. This archive includes, of course, the executive order enacting the SOM as well as the publicly available records of the AEQ board meetings. Additionally, I examined the Law of Education Quality Assurance (LEQA), the report of the Presidential Advisory Commission of 2006, the reports issued by the Education Commission in Congress to inform plenary sessions about reform bills, the minutes of the plenary sessions when these bills were addressed, the session minutes of the National Council of Education (NCE), and two key policy documents: the report by the "SOM Commission" submitted to the AEQ and NCE for consideration, and

the report by the Taskforce created in 2014 for the revision of the test-based accountability regime. These texts explicitly discuss accountability policy and address the techniques and procedures that were eventually adopted. I examined all these materials for the use of moral judgments (claims of fairness or unfairness) to justify policy positions regarding school assessment methods.

Likewise, to make sense of the positions of the experts directly implicated in the SOM's design and implementation, I researched who these experts were, making sense of their institutional affiliations, professional backgrounds, and, importantly, their political leanings. I inferred the latter from publicly available information (e.g., blogs, newspapers, websites) about their ties to the private or public sector of education provision; their professional involvement in academic institutions and think tanks publicly known for their conservative or progressive tendencies; and their prior participation in government bodies appointed by conservative or progressive presidents.

I supplemented this archival work with 57 in-depth interviews I conducted in Chile in the summer of 2014 and the fall of 2015. Interviewees were selected through strategic and snowball sampling techniques [Maxwell 2012], and included different stakeholders involved in the controversies over the regulation of education markets in Chile. Many of the interviewees were experts who hold (or have held) key positions in the governance structures of education, in think tanks involved in education policy networks, or in leading educational research institutes. Among other key informants, I interviewed the former CEO of the AEQ, the president of the AEQ's board, the president of the NCE, and the experts in charge of writing some of the hallmark reports and policy documents leading to the implementation and subsequent revision of the quality assurance system. Although the empirical material cited in this paper comes almost exclusively from my archival analysis, many of the conceptual issues, questions, and information that helped me reconstruct the SOM's genealogy came out of conversations with my informants and interview respondents.

The SOM in the Making

A stylized genealogy of a school accountability device

The education authorities of Chile set out to develop high-stakes accountability regulation in the context of broader transformations in

the political economy of school reform in Chile. The decisive shift began in 2006, when hundreds of thousands of high-school students took to the streets and occupied school buildings across the country [Donoso 2013]. Declaring *quality* education a universal right rather than an exclusive market good, the student movement condemned profit and discriminatory enrollment in the voucher-funded private sector, and demanded the state reassume direct control and funding of public schools. The students' manifesto⁷ also claimed that education privatization was furthered by the public use of school rankings based on average proficiency in the SIMCE (*Sistema de Medición de la Calidad Educativa*), a standardized test explicitly created by the Chicago Boys during Pinochet's dictatorship to inform parental choice in the context of privatization and decentralization [Benveniste 2002].

The protests provoked intense intra-elite controversy. While progressive reformers sought to give preferential treatment to public schools and set restrictive barriers for private voucher schools to operate in the market, conservatives contended that the state ought to remain impartial regarding the type of provision and focus merely on overseeing and enforcing quality standards [Falabella 2015; Larroulet and Montt 2010]. After two years of political stalemate, and confronted with the powerful opposition of an alliance of private providers, mainstream media, and influential think tanks against state interference [Corbalán Pössel and Corbalán Carrera 2012], the progressive government of Michelle Bachelet crafted a bipartisan compromise that paved the way for reform but alienated the social movement that had demanded it.⁸ The consensus was that market failures were to be "repaired" to assure quality education for all.

Political conflict over the regulation of the k-12 education market gave rise not only to the LEQA, but also the SOM itself, with its mandate for fair responsabilization. The urgency of writing explicit safeguards into the law became apparent in the wake of the triumph, for the first time since the democratic transition, of a right-wing coalition in the 2009 Presidential election. Even as the new government reaffirmed its support to the LEQA—still making its way through Congress—to honor bipartisan accords, it implacably insisted on holding failing schools

⁷ See *Resoluciones I Congreso Nacional de Estudiantes Secundarios* at http://www.opech.cl/bibliografico/doc_movest/resoluciones_congreso_nacional_de_estudiantes_secundarios.pdf, retrieved on January 19th, 2017 [author's translation].

⁸ See the terms of the compromise incorporated in the History of the General Law of Education at <http://www.bcn.cl/historiadelaley/nc/historia-de-la-ley/4713/>, retrieved on January 19th, 2017.

accountable to sovereign consumers. In June 2010, the Ministry of Education (Mineduc), directed by Chicago-trained economist Joaquín Lavín, released the last results of the SIMCE (which measures proficiency) and mailed them to every family in Chile along with a map in which all the schools of the district were colored based on average test-scores in red, yellow, or green, like a stoplight. “Lavín’s stoplights”—*los semáforos de Lavín*, as they became popularly known—further validated the market-fundamentalist claim that voucher-funded private schools generally outperform municipal schools. For anti-market advocates, the maps showcased quite the contrary: the extreme overlapping of educational opportunity gaps with spatial patterns of class-based segregation⁹. So powerfully did SES predict test scores that some of the poorest districts in dense metropolitan areas did not have a single school in yellow or green¹⁰. Anti-market critics pointed to the tautology of performance measures that merely reified pro-market reformers’ trashing of public schools serving mostly poor students.

Public controversy over the implacable application of market discipline to drive school improvement certainly resonated within the still-unfinished legislative debate. This final round of deliberation in Congress remained plagued by skirmishes over the use of “raw” test-scores—that is, percentages of students attaining *proficiency*—as the yardstick of school quality; petitions to compute adjustments by SES or even incorporate value added measures to capture *effectiveness*—the specific effect of instruction on student growth; and suspicion of the distortions these statistical manipulations might introduce to the universal enforcement of common core standards. Nevertheless, lawmakers passed the LEQA in early 2011, right before the outbreak of a new wave of student protests further radicalized the conflict over the market model of education, monopolizing the public agenda for the rest of the year.

Simulating commensuration

It was in this polarized atmosphere that the Mineduc and the newly created AEQ undertook the tricky task of developing a “fair” and “socially sensitive” methodology to commensurate schools with a single

⁹ See <http://ciperchile.cl/2011/03/03/las-desigualdades-sociales-que-desnudo-el-control-vertido-semaforo-del-mapa-simce>, retrieved on January 19th, 2017.

¹⁰ Several of my interviewees recognized the release of these maps as a very controversial

policy, for it illustrated quite well the government’s extreme pro-market ideology. Lavín was removed in 2011, at the peak of the student protests of that year.

metric of quality. Already in late 2010, upon the imminent enactment of the LEQA, the government set up an ad hoc commission of specialists and tasked it with proposing a tentative SOM to the soon-to-be AEQ. This “Ordinalization Commission” was headed by conservative economists in charge of the SIMCE administration and other divisions of the Mineduc, and worked in consultation with another economist hired by the United Nations Development Program but institutionally affiliated with a private university well known for its conservative leanings (see figure 1). Over the next three years, the commission reviewed the international literature and surveyed national systems of high-stakes accountability. It also gathered and consolidated all the necessary data sources, revised and calculated all the necessary indicators, and, more crucially, conducted simulations based on alternative model specifications and estimation methods. To validate and review drafts, the commission convened three rounds of meetings with school administrators and principals and organized twelve workshops with independent experts (mainly economists, psychometricians, educational psychologists, and sociologists). In total, the commission consulted with more than 140 people.

The work of the Ordinalization Commission was documented in detail in the report *Fundamentos Metodología de Ordenación de Establecimientos* (Foundations of the School Ordinalization Methodology). The bulk of the language and formulae of the 2014’s Executive Order came out of this report.¹¹ After recognizing *quality* and *equity* education as a political mandate stemming from the student movement of 2006 and then made official by the Presidential Council, as well as the subsequent bipartisan consensus on preserving the voucher-funded public-private scheme of service delivery, the report made it clear that the proposed design of the SOM sought to fulfill simultaneously two different requirements: the “fair responsabilization” of schools, and the achievement of “learning standards” by students [39].

This attempt at reconciling in a single instrument competing claims for valuing school effectiveness *and* enforcing proficiency benchmarks severely constrained the commission’s decisions and justifications. But it also facilitated the translation of the moral conundrum (proficiency vs. effectiveness) into a technical matter concerning the identification of the “correct” econometric equation.

¹¹ The 575-page document also includes a detailed account of the elaboration process as well as the statistical simulations discussed in this paper. I especially thank Francisca

Dussailant, Luz María Budge, and Carlos Henríquez for giving me access to this crucial data source.

THE MORAL LIFE OF ECONOMETRIC EQUATIONS

FIGURE 1
Experts' backgrounds, affiliations, and partisan leanings

| | Category | Academic Background | Institutional Affiliations | Other Affiliations | Partisan Leaning |
|------------|--|---|--|--|------------------|
| Expert #1 | Experts in charge of drafting the SOM proposal | Commercial Engineer, Economics & Public Policy - University of Chile | CEO of AEQ 2012-2014, Superintendent of Education 2018 to present | Ministry of Finance, Ministry of Economy, Mineduc, Adolfo Ibañez Univ, CEP Think Tank | Conservative |
| Expert #2 | | Educational Psychology - Catholic University of Chile | Director of Research at Mineduc's Curriculum and Evaluation Unit | CEP Think Tank, Voucher-Funded Private Sector | Conservative |
| Expert #3 | | PhD in Economics - Catholic University of Chile, MA Education - UNC Chapel Hill, Civil Engineer - Catholic University of Chile | UNDP expert in charge of SOM Report | University of Development, NCE member 2013-2018, Minister of Finance 2018 - | Conservative |
| Expert #4 | Regulators with decision-making authority over the SOM | English Education - Catholic University of Chile | AEQ Board President 2011-2014, 2018 to present | Dean of Education Finis Terrae University and University of the Andes, Voucher-Funded Private Sector, Liberty and Development Think Tank | Conservative |
| Expert #5 | | Educational Psychology- Catholic University of Chile, MA Social Policy - London School of Economics | AEQ Board Member | Voucher-Funded Private Sector | Conservative |
| Expert #6 | | Computer Engineer (Educational Testing and Technology) - Catholic University of Chile, MA Computer Science - Georgia Tech, PhD Technical Sciences - Eidgenössische Hochschule of Zürich | AEQ Board Member | School of Engineering, Catholic University of Chile | N/A |
| Expert #7 | | MA Sociology - Catholic University of Chile | AEQ Board Member | School of Education, Diego Portales University Fundación Chile, Democratic Revolution Political Party | Progressive |
| Expert #8 | | Psychologist and MA Communication and Education - Catholic University of Chile | AEQ Board Member, AEQ Board President 2014-2018 | Educational Services Industry, Christian Democratic Party, Diego Portales University, World Bank, Inter-American Dev. Bank | Progressive |
| Expert #9 | Independent advisers of the AEQ | Psychology - Catholic University of Chile, PhD UCLA | Professor at the Catholic University of Chile - Independent expert hired by the AEQ to comment on SOM design | University-affiliated educational testing services | Progressive |
| Expert #10 | | Commercial Engineer & Economics - University of Chile, PhD Economics - UCLA | Professor at the Catholic University of Chile - Independent expert hired by the AEQ to comment on SOM design | School of Economics - University of Chile | Conservative |

Source: Author's elaboration based on *Fundamentos Metodología de Ordenación de Establecimientos*, prepared by the Ordinalization Commission and the Research Division of the Agency of Education Quality 2013.

The statistical aspects of this controversy are at least puzzling, if not downright confusing. Yet they are central to the school valuation problem at stake. That the school performance categories stipulated

in the LEQA were drawn from a composite index generated technical consensus among experts. Experts similarly agreed that index scores overwhelmingly reflected schools' average proficiency in common core standards, measured by routine standardized tests.¹² Experts further concurred that, through statistical modeling and adjustment, *absolute* proficiency levels could be rendered *relative* to individual- and school-level data that roughly approximated student SES, thereby honoring the principle of fair responsabilization. Finally, the alleged requirement of balancing presumably incompatible quality standards with a single ordinal measurement justified a second correction: the introduction of proficiency “thresholds” that, once the SES adjustment had been computed, automatically reassign very high or very low performing schools (in terms of absolute test scores) to the upper and lowest categories of the ordinalization—a sort of automatic exemption from the effectiveness standard.

The most crucial decision was which regression equation should be used to estimate the index and then perform balanced post-estimation recalibrations. To aid decision-making, the commission simulated and compared the predicted distribution of five alternative SOMs, each corresponding to a different modeling strategy: Ordinary Least Squares (OLS), Fixed Effects model (FE), and three different specifications of Hierarchical Linear Models (HLM). As Figures 2 and 3 show, the weight of SES adjustments significantly varied according to the estimation method and model specification. While FE and HLM-1 brought the SOM closer to the unadjusted index of absolute test-scores, HLM-2 and HLM-3 returned adjusted rankings that greatly deviated from proficiency scores. OLS fell in the middle, allowing for moderate, Solomonic adjustments.

Although the report never made it explicit, this gap in levels of SES adjustment depended on the degree to which the *contextual* effects of SES (that is, at the school-level) were made endogenous to the model and estimated separately from the *individual* effects of SES (at the student level). This contextual effect—namely, the impact of the overall context of poverty or wealth in which a school operates—was better captured by HLM-2 and HLM-3 when key students' socioeconomic characteristics

¹² The Law stipulates that at least two thirds of the index scores must reflect students' proficiency levels in common core standards, the other third corresponding to measures of “student growth” and “other indicators of quality” (taken from surveys and administrative

records). Because between-school variation in these other indicators is considerably lower than in test scores, proficiency weighted considerably more than two-thirds in the final distribution of the unadjusted index.

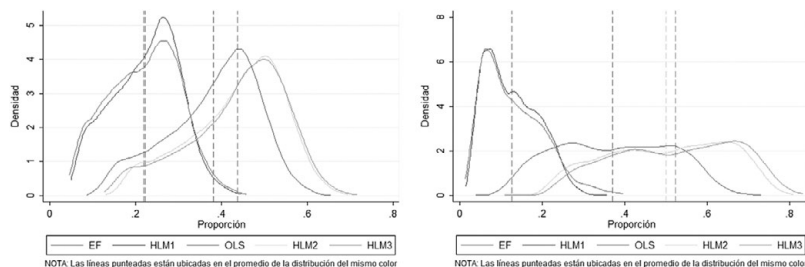
FIGURE 2
Characteristics of the models considered to compute SES adjustments

| Model | Student-Level Input Variables | School-Level Input Variables | School Quality Identification | Coincidence with Absolute Test-Scores (Proficiency) |
|--|--|---|--|--|
| <i>Ordinary Least Squares (OLS)</i> | Vulnerability Mother's education School recent enrolment of high and low achievers Indigenous background School is in rural area School District Crime Rate | | School average of adjusted (student-level) residuals | 65% (elementary) 72% (secondary) |
| <i>Fixed Effects (FE)</i> | Vulnerability Mother's education School recent enrolment of high and low achievers Indigenous background School is in rural area District Crime Rate School ID (dummy) | | School average of adjusted (student-level) fixed effects residuals | 76% (elementary) 87% (secondary) |
| <i>Hierarchical Linear Model (HLM-1)</i> | Vulnerability Mother's education School recent enrolment of high and low achievers Indigenous background School is in rural area School District Crime Rate | School is in rural area | Adjusted (school-level) random effect component | 79% (elementary) 89% (secondary) |
| <i>Hierarchical Linear Model (HLM-2)</i> | Vulnerability Mother's education School recent enrolment of high and low achievers Indigenous background School is in rural area School District Crime Rate | Mother's education (averaged) School is in rural area | Adjusted (school-level) random effect component | 57% (elementary) 51% (secondary) |
| <i>Hierarchical Linear Model (HLM-3)</i> | Vulnerability Mother's education School recent enrolment of high and low achievers Indigenous background School is in rural area School District Crime Rate | Vulnerability (averaged) Mother's education (averaged) School recent enrolment of high and low achievers Indigenous background (averaged) School is in rural area School District Crime Rate | Adjusted (school-level) random effect component | 54% (elementary) 47% (secondary) |

Comment: The main differences between HLM-2 and OLS (the two models subjected to a vote in the final round) is the inclusion of SES indicators averaged at the school level in HLM-2. In OLS, school quality scores are computed by averaging adjusted residuals for each school, while in HLM school quality scores derive from adjusted random effects at the school level.

Source: Author's elaboration based on *Fundamentos Metodología de Ordenación de Establecimientos*, prepared by the Ordinalization Commission and the Research Division of the Agency of Education Quality 2013.

FIGURE 3
 Simulations of weights of the SES adjustment produced
 by the Ordinalization Commission



Comment: This can be interpreted as a measure of the degree to which each model modifies the original distribution of average test scores (proficiency) in elementary schools (left) and secondary schools (right). Note that EF and HLM-1 yield the smallest adjustments, HLM-2 and HLM-3 the highest, and OLS falls somewhat in the middle.

Source: *Fundamentos Metodología de Ordenación de Establecimientos*, prepared by the Ordinalization Commission and the Research Division of the Agency of Education Quality 2013.

such as mother's education or vulnerability were averaged and added back in at the organizational level. In class-segregated education systems (i.e., large between-school SES variation) such as Chile's, the contextual effect of SES is usually large, so its parametrization and subsequent inclusion in post-estimation adjustments significantly reduces the remaining between-school variance—in this case, the variability of school quality, thus defined. Hence the contrast with the other models was sharp. Although HLM-1 and the FE model did consider the clustering of students into schools, their (mis)specification of school effects simply ignored the contextual dimension of SES. The OLS regression, on the other hand, directly assumed no nesting of students within schools, so the model only partly captured the between-school SES effect but confounded it with the within-school SES effect.

Model choice, and the weight of SES adjustment the decision carries with it, therefore appeared highly consequential for determining the ultimate prevalence of proficiency vis-à-vis effectiveness in the predicted ordinalization, something the report made crystal clear:

The higher β , the higher the adjustment (SES index) and less will be required from the most vulnerable vis-à-vis non-vulnerable schools in the Unadjusted Index [that is, absolute proficiency] in order to move up in the Ordinalization categories.

Then, if the adjustment level is high, vulnerable schools may achieve high values in the Final Index with low scores in the Unadjusted Index [486].

The interpretation of the distributional consequences associated with alternative calibrations of the instrument indicates that fear of *effectiveness* outweighing *proficiency* standards prevailed over any other consideration. After sorting the five models by SES effect size and discarding HLM-1 and FE for yielding meager adjustments, the report noted that in HLM-2 and HLM-3 “there are schools that are required up to 60 points less in terms of the Unadjusted Index [a measure of proficiency scores] to move from one Ordinalization category to another” [487]. In these models, “the adjustment level is such that a considerable number of schools with similar test scores will be classified in three different categories” [487].

Most unacceptable to the Ordinalization Commission was the degrading of prestigious elite schools, popularly known in media rankings for their high level of proficiency. In this respect, the report warned that choosing HLM-2 or HLM-3 would mean that “nearly all the least vulnerable schools with the best test scores of the country classify into the Medium Performance”, which “explains why these models require the disproportionate use of [post-adjustment proficiency] thresholds” [490]. Finally, the report contended that analyses of model stability proved HLM-2 or HLM-3 to be the most sensitive to cross-time variation in the number of schools classified at the bottom of the ordinalization. For the commission, adopting such a time-sensitive model meant compromising the high-stakes purposes built into the design of the policy device. For “it will be more unlikely for a school to be systematically classified into the Insufficient Performance category for five years” [515]—the legal requirement to make a school subject to intervention and closure.

Thus construed, the distributional consequences associated with different alternative models made the Ordinalization Commission’s choice of model obvious: OLS, deemed the model that better achieves an equilibrium between too-much and too-little, was preferable. In this model “the weight of SES adjustment is more similar to that of the models that correct the most, and it does not yield counterintuitive results”; per the Commission, counterintuitive results were those that degraded high-proficiency and boosted low-proficiency schools. Therefore, OLS “achieves a better balance between the two objectives of the SOM: to hold schools responsible with fairness and create equal expectations for all the students in the country” [517].

Arbitrating between models' moral values

In October 2012, the board¹³ of the AEQ began holding official sessions. The first order of business was to deliberate and decide on the proposal drafted by the Ordinalization Commission. To make a final judgment on the suggested statistical model, the board solicited the opinion of two independent consultants¹⁴. The first of these experts was generally sympathetic to the commission's suggestions. He acknowledged OLS's weaknesses in terms of endogeneity and misspecification, problems that could be easily fixed if value-added models were used. However, the latter would require panel data, currently unavailable. And anyway—he went on—value-added models would not resolve all these problems. Likewise, for this specialist, hierarchical linear models could offer a reasonable alternative to handling selection biases, but these models “do not necessarily dominate over OLS” [550]. All in all, he fundamentally agreed on the two criteria followed by the commission: classifying according to whether predicted school residuals fall below or above the line produced by regressing test scores to SES, but also reclassifying when schools reach minimum and maximum levels of absolute proficiency.

The second expert, however, was less sympathetic. He expressed support for adjusting the index by factors beyond school control, which “acknowledges the well-known fact that achieving a certain level of quality education is harder or easier depending on conditions such as the ones this proposal considers” [534]. In this regard, he welcomed that the SOM “seeks to represent in fairer fashion what is possible to achieve in different educational context” [534]. And yet, the choice of statistical model concerned him greatly. For him, OLS's assumptions on the distribution of the error term were simply untenable, which raised serious validity issues concerning what in applied economics is known as “identification strategy”—the procedures adopted to assert that the estimated parameters approximate the real parameters as estimated under experimental conditions. Using HLM, instead, would have at least allowed for a better treatment of residuals to identify the parameter of interest, namely, school quality:

¹³ At the AEQ's board seat five technocrats with recognized trajectory in educational affairs, appointed by the President every three years in tandem of two and three. Appointments must follow a “plurality criterion” to reflect diversity of policy perspectives.

Because the AEQ started operation during Piñera's conservative government, conservative members held a majority of seats. Its first CEO was also appointed by the president following a public tendering process.

¹⁴ See session minutes Apr-03-2013.

To hold schools responsible for their outcomes, it is necessary to make sure that the utilized analysis model clearly identifies that portion of the examined variability that actually depends on the school and not on other factors. In this sense, the use of residuals of a conventional regression equation (with student-level data), as baseline to establish school effects, in practice confounds the part attributable to the school with another that depends on the student-level error. The use of HLM allows for separating these components [...] These models identify a school-level residual that is different from the student-level residual, which provides better statistical support to assert that a particular residual represents the school effect once all the control variables are included (the proposed method averages the individual residuals of all students in the school as expression of the school effect, after having run a regression at the individual level). [pages 535-536, 540-541]

The question at stake, again, was how much of the SES contextual effect was being disguised as school quality, supposedly captured by the residual part of the model. Underlying the apparently esoteric technical quandary about statistical artifacts was a serious first-order disagreement [Abend 2014]: one over the extent to which the four-tiered ranking of quality values performed by the soon-to-be ordinalization device might potentially deviate from the status hierarchies already built into the highly segmented Chilean school market [see e.g., Podolny 1993]. Simply put, should absolute proficiency, better reflected by OLS post-estimation adjustments, prevail as the valuation criterion, the resulting hierarchy of school quality would tacitly validate the class-based tiers in which the market is organized (see Table 1). Should the valuation device be more sensitive to the specific school effect on proficiency—as in HLM-2 and HLM-3 post-estimation adjustments—the performed ordinalization might potentially invert, or at least counter, the market's status order¹⁵.

In light of this discrepancy, opinions split between board members sympathetic to the conservative government, and those gravitating towards the progressive opposition (Figure 1). They did not, however, openly and explicitly confront this ideological disagreement. Instead, board members, contrasting the statistical simulations based on HLM-2 and OLS—that is, comparing the orders of worth performed by this or that calibration strategy—, resorted to moral judgements to justify their technical decisions for this or that econometric equation.

¹⁵ Muñoz-Chereau and Thomas [2016] estimated contextualized value-added models for Chilean secondary education and found school practices account for only 13.2% of the between-school variation in test scores. Also, San Martín and Carrasco [2012] found that a 60% of Chilean schools change their

effectiveness status if assessed by value added relative to unadjusted test scores. Moreover, 48.3% of public schools would improve their performance category if reclassified using value added, while 41.4% of private voucher schools would be degraded—Pure political dynamite!

TABLE 1
Mean and percentile distribution of within-school % of socioeconomically vulnerable students by school sector

| Percentile | Municipal Public | Voucher-Private | Private | Total |
|------------|------------------|-----------------|---------|-------|
| 10 per | 35.4 | 16.3 | 0.0 | 19.1 |
| 25 per | 43.1 | 26.9 | 0.6 | 34.7 |
| 50 per | 51.1 | 39.9 | 1.9 | 47.0 |
| 75 per | 59.8 | 53.6 | 3.9 | 57.1 |
| 90 per | 69.0 | 65.8 | 8.0 | 67.5 |
| Mean | 51.5 | 40.6 | 3.5 | 45.0 |

Comment: Municipal public schools have a higher % of vulnerability than voucher private schools, which in turn have a higher % of vulnerability than non-voucher private schools. The higher the SES adjustment, the higher the boost in quality scores for municipal public schools.

Source: Author's elaboration based on statistics reported in the document *Fundamentos Metodología de Ordenación de Establecimientos*, prepared by the Ordinalization Commission and the Research Division of the Agency of Education Quality 2013.

What did technocrats make of the comparison of ante- and post-adjustment simulations based on both estimation methods? (See Table 2 for a small sample of this comparison.) A minority of board members voiced their preference for HLM-2, under the explicit argument that greater adjustments by SES better captured school effectiveness¹⁶. School quality, they contended, could only be gauged relative to school practices. A “fair responsabilization” of schools for the service they provide must value schools for effects they can reasonably be assumed to cause. And by no reasonable standard, they claimed, should the responsabilization of schools in disadvantaged communities depend on their students' social class background.

The board's majority, however, opted for OLS. Echoing the market-fundamentalist perversity thesis [Hirschman 1991; Somers and Block 2005] that providing unconditional support to the poor disincentivizes work and exacerbates dependency, one board member warned that larger adjustments by SES would significantly alter low-SES public secondary schools' rank in the final ordinalization (i.e., move them up two positions, from “insufficient” to “satisfactory”), which might constitute a

¹⁶ Session minutes of Apr-08-2013.

THE MORAL LIFE OF ECONOMETRIC EQUATIONS

TABLE 2
*Some characteristics of the simulated ordinalizations based on OLS
 (model choice) and HLM-2 (alternative)*

| | HLM-2 | OLS |
|--|-------|-----|
| Vulnerable/non-vulnerable schools' gap in test scores necessary to change category (elementary) | 60 | 45 |
| % weight of SES adjustment (elementary) | 45 | 38 |
| % weight of SES adjustment (secondary) | 51 | 37 |
| % of schools in High Performance category with proficiency below schools in Insufficient Performance category (elementary) | 20 | 2 |
| % of students below proficiency in High Performance schools (elementary) | 24 | 18 |
| % of schools to be reclassified by absolute proficiency thresholds (elementary) | 15 | 4 |

Source: Author's elaboration based on statistics reported in the document *Fundamentos Metodología de Ordenación de Establecimientos*, prepared by the Ordinalization Commission and the Research Division of the Agency of Education Quality 2013.

disincentive for failing schools (in terms of absolute levels of proficiency) to seek improvement.

Correcting [by SES] does justice [to these schools] because it takes into account the hardships these schools face, but on the other hand, this does not necessarily provide the proper signal for the system to improve and take charge of students' [learning] difficulties. For that reason I choose the model that adjusts to a least extent, knowing that only value added models really allow for a fair classification of schools.

This perspective was seconded by the president of the board, who argued that smaller adjustments by SES prudently prevented too much distortion of gross results, thereby taking account of “vulnerable schools' greater learning hardships.” Endorsing the president's position, a third member deemed HLM-2 adjustments “unfair” to the poor, since the same (absolute) score would categorize well-off schools as “unsatisfactory” quality and vulnerable ones as “high” quality; a situation that would “crystalize too-low expectations [for the poor] and result in inadequate targeting of state resources”—again, the perversity thesis.

The argument for adopting a conservative approach to SES adjustment was then made explicit in a policy brief¹⁷ detailing the School

¹⁷ See <https://slidedoc.es/minuta-informe-metodologia-de-clasificacion-en-categorias-de-desempeno-pdf>, retrieved on September 15th, 2017.

Classification Methodology (note the adoption of the new terminology) released by the AEQ:

Some HLM specifications generate very different achievement expectations between students from different SES contexts, which would go against the principle of quality and equity in education provided by art. 3 of the General Education Law. In practice, using a method that generates very different expectations implies giving a signal that would consider low-SES students' results as excellent, when those same results would be considered unacceptable for students in more favorable contexts. In other words, the classification generated by some... methods risks masking the inequities in the educational opportunities that schools provide in our country [7-8].

And beyond the performative effects that higher SES adjustments might have on expectations, the AEQ also alleged pragmatic reasons concerning the correct targeting of resources to those who actually need them. As the law mandated that schools classified at the bottom automatically qualify for government supervision and guidance, assigning low-quality status to a larger number of poor schools yielded a "fairer" pattern of resource allocation. Adjusting by HLM-2, however, "would fail in the correct identification of schools that need support", which would "deprive underperforming low-SES schools from the chance of receiving Orientation and Evaluation Visits and, then, Technical Pedagogical Support, benefiting those schools from less vulnerable contexts" [8].

To base these arguments on evidence, the AEQ presented simulation results of pre- and post-adjustment ordinalizations for different model selections, indicating that adjusted HLMs "with averaged variables" yielded "classifications in which more than 20% of schools in the category High Performance (schools with students relatively more vulnerable), would have worse learning results [in terms of absolute SIMCE scores] than schools with less vulnerable students in the category Insufficient Performance" [14].

Sanctifying common sense

In December 2013, a second regulatory body, the National Council of Education (NCE), released, in accord with the LEQA, a technical report¹⁸ on the SOM design proposed by the Ordinalization Commission and approved by the AEQ. The report was based on a study conducted by eight researchers (five national, three international), who evaluated whether the valuation device followed the criterion that "the resulting

¹⁸ See <https://www.cned.cl/resolucion-de-acuerdo/informe-metodologia-de-ordenacion-de-establecimientos>, retrieved on September 15th, 2017.

categories be fair according to the diverse (socioeconomic) contexts, and that they favor mobility [of poorly performing schools] to superior levels” [1]. Overall, the NCE celebrated the careful and responsible work done by the AEQ, and suggested a smooth, cautious implementation of the new classification system, in order to protect its legitimacy. The NCE further acknowledged the technical challenges that constructing these measurements represent in terms of internationally validated scientific standards, which implied “a deep knowledge of psychometric and econometric topics” [6]. The NCE finally stated that it was acutely aware that the SOM’s legitimacy depended not only on technical aspects but also on its appropriation and implementation by users, and on its consequences.

In terms of how the SOM understood quality, the NCE applauded the inclusion of a conception of equity as a key component of quality. It also praised the methodological consideration of definitions of quality commonly used by the empirical literature on effective schools. More or less explicitly, however, the NCE validated a notion of quality reduced to schools’ ability to “produce” students who “overcome Insufficient Learning Levels” [7]. In other words, quality amounted to schools’ meeting cognitive learning benchmarks in specific areas of knowledge, that is, proficiency. Allegedly an inherent component of quality, equity meant nothing but the achievement of such basic proficiency standards by all.

In light of this conception of quality, it is no coincidence that the report approved of the econometric model finally adopted to compute the scores used in the classification. After acknowledging the lack of academic consensus on the best method to separate school and class effects on student learning, the NCE noticed that the AEQ’s decisions were empirically grounded in “an analysis of the obtained classification;” simulations that considered “different educational realities to an important degree” [8]. Hence the NCE endorsed the AEQ in its intent to produce estimates that guarantee that “schools be held responsible [after] controlling for those associated factors that are out of their intervention possibilities” [8]. On the other hand, however, the NCE’s report agreed with the AEQ’s criteria of preempting “different learning expectations for students from different SES backgrounds” [8]. These provisions, ultimately, were necessary “to guarantee a fair classification” [8].

Afterlife: adjusting the adjustment

Controversy subsided, but only briefly. The enactment of the SOM not only implied that public categorizations of school quality would now be

rendered official and legally binding; it also entailed the intensification and expansion of mandatory testing to more grades and curricular areas. This was made explicit in the Testing Plan of 2012¹⁹, wherein the Mineduc predicated the increase of national tests from nine to seventeen in 2015 upon the adequate implementation of the SOM. For the conservative government, making testing ubiquitous was necessary “to carry out a valid and reliable ordinalization of elementary and secondary schools” [7]. Achieving robust measures of school effectiveness amounted to a moral imperative, the fulfillment of which was “fundamental to contributing to a fairer and more equitable ordinalization of schools” [10].

This intensification and expansion of high-stakes testing was the context of *Alto al SIMCE* (Stop the SIMCE), a public campaign “in rejection of the standardization rationality and the effects of the measurement and quality assurance system.”²⁰ Launched in mid-2013—right before the Presidential election—by a small group of young educational researchers, *Alto al SIMCE* advocated for “a form of evaluation in concord with an education understood as a social right and not as a commodity.” The campaign soon won the endorsement of hundreds of renowned educational researchers, artists, and intellectuals. It also enlisted the support of the teachers’ union, the main student federations, as well as their popularly acclaimed former leaders who were now running for Congress.

In March 2014, a new center-left coalition rose to power with the explicit programmatic commitment of rolling back market-oriented policies in education, including a revision of the testing regime. The new authorities of Mineduc commissioned a group of education specialists of diverse political backgrounds to produce a report²¹ on possible fixes to accountability policy: the so-called “taskforce for SIMCE revision.” *Alto al SIMCE* was among the myriad stakeholders invited to the taskforce’s hearings. There the campaign proposed to suspend national testing for three years, stop the implementation of the SOM, and smoothly transition to a new evaluation system based exclusively on survey tests, complemented with formative assessments at the district and school level [28].

¹⁹ See http://www.agenciaeducacion.cl/wp-content/uploads/2013/02/Plan_de_Evaluaciones_Actualizado.pdf, retrieved on September 15th, 2017.

²⁰ See <http://www.altoalsimce.org>, retrieved

on September 15th, 2017.

²¹ See <https://www.mineduc.cl/wp-content/uploads/sites/19/2015/11/Informe-Equipo-de-Tarea-Revisi%C3%B3n-Simce.pdf>, retrieved on September 15th, 2017.

Although this proposal received the endorsement of a group of seventeen lawmakers, most of them belonging to the government coalition²², only a minority of the taskforce, linked to the student movement and the teachers' union, and more sympathetic with *Alto al SIMCE*, firmly pushed for a significant rollback. This group of experts recommended a two-thirds reduction in annual testing [86-88]. They also argued, straightforwardly, for the suppression of the SOM, and its replacement with a mechanism of targeted assistance for those schools that fall below a proficiency threshold, measured by unadjusted test scores.

The majority report adopted much of the revisionist language, but it approved of moving forward with the pilot implementation of the SOM, highlighting that at least the new ordinalization “takes students’ socioeconomic characteristics into account” [43]. The SOM, therefore, represented “progress relative to communicating only average scores in SIMCE, which induces erroneous interpretations of school effectiveness” [43]. And still, the majority report recognized multiple validity issues regarding the design of the SOM. The taskforce questioned the application of proficiency “thresholds” at the two bounds of the post-adjustment index distribution, which assign “schools with high SIMCE in[to] the highest category, even if they have not been effective when considering their SES.” Thresholds also “leave in the lowest category those schools with very low absolute results, even if they achieve better scores than other schools in the same SES level” [62]. Thus, in what appeared to be a re-instantiation of the moral conundrum pervading the design of the SOM, the report alluded to the dangers of schools being classified in the extreme categories according to two different criteria: their “absolute results” (proficiency), or “their effectiveness” (value added).

As I noted in the previous section, ambiguity of criteria stemmed from the intricate coexistence of apparently incommensurable perspectives on school accountability. Yet, this time experts’ adjudication was crystal clear: using proficiency thresholds in the SOM restricted, if not directly contradicted, the SES adjustment mandate that “seeks to give weight to factors attributable to schools and approximate to an evaluation of their effectiveness as to hold them responsible in a fair way and reflect their performance” [62]. On the closely related but more “technical” matter of whether the estimation method chosen to compute adjustments fairly honored such a mandate, the taskforce made no comments at all.

²² See http://www.altoalsimce.org/wp-content/uploads/2015/09/Carta_Parlamentarios_

[Cambios_Legales_al_SIMCE-Dic-2014.pdf](#), retrieved on September 15th, 2017.

In October 2019, amid the largest social outbreak in Chilean history, and upon the imminent release of the fourth legally binding ordinalization, the Senate Education Committee began discussing a bill introduced by progressive lawmakers that derogates the LEQA's mandate to close every school classified four consecutive years in the Insufficient category. In their preamble²³, the proponents of the bill review the international evidence of the negative consequences of school closures and warn about the pernicious logic of standardized testing, denouncing that public schools, not surprisingly, represent three out of four schools at risk of closure. Opponents of the bill, in turn, hotly defended the punishments and rewards associated with the SOM, arguing that adjustment by SES prevent socioeconomically different schools from being commensurate with the same metric. The indictment or defense of high-stakes accountability notwithstanding, lawmakers offered no comment on the technical validity of the statistical method used to determine which school should be deemed underperforming. The issue would remain absent throughout the committee's hearings and the plenary session that, in August 2020, approved the bill by a slim majority and sent it to the House. The experts, it seems, had already adjudicated.

Discussion

The long road that goes from the politicization of market-based inequality in education, to the adoption of accountability mechanisms that intentionally factored between-school SES variation into assessments of school quality, reveals that a moral economy was constantly at play in the design of the SOM. Policymakers' quest for fairness in putatively objective, impersonal assessments of school quality required forming subjective, personal judgment on the potential performative effects associated with different adjustments of the new commensuration instrument. The justificatory rationales at play in this controversy reflected an intricate moral economy, intelligible only in light of the anti-market versus pro-market cleavage that crosscuts the Chilean polity regarding the administration and provision of social services. The clash between two notions of "quality" operationalizing two distinct conceptions of "fairness" became inevitable. And yet, experts' translation of a

²³ See Education Committee Report (Bulletin Nr. 12.980-04) <https://www.senado.cl/appsenado/index.php?mo=sesionessala&ac=getCuenta&iddocto=64215#>, retrieved on January 22nd, 2020.

priori antithetical conceptions of market regulation and accountability into competing claims for fairness that, nevertheless, were both rooted in shared ways of going about settling their moral disputes, helped bridge the ideological gap. Reducing this disagreement to a tradeoff between proficiency and effectiveness that could be technically handled through econometric adjustments and post-estimation thresholds blurred the ideological divide at stake. Probably the SOM's most consequential "performativity effect" consisted of rendering commensurable alternative moral ways of making the SOM "performative."

Eventually, despite warnings and appeals to fair responsabilization by progressive experts, the conservative method of SES adjustment prevailed, better suiting the standpoint of market-minded policymakers. A minority of progressive experts equated quality to measures of school effectiveness. Invoking fairness in relation to providers (especially poor schools) for the sake of a just distribution of responsibility for educating disadvantaged students, primarily served by public institutions, they advocated for large adjustments by SES. Yet the majority of conservative experts, unsuspecting of proficiency rankings, especially jealous of preserving the SOM's efficacy as an incentive to organizational improvement, and primarily concerned with giving families (poor families in particular) the correct market signals to exercise choice, advocated for moderate adjustments by SES. Because they conceived fairness on behalf of consumers, they understood quality in reference to achievement of proficiency benchmarks, irrespective of the school's relative contribution to that outcome. The official narrative justifying this decision was that, beyond the scientific convention that evaluations of school quality should rely on unbiased estimates of school effects on student progress, ordinalizations based on greater adjustments by SES might lead to policy inefficient, if not politically illegitimate, results. Models that (at least for some) produced less biased estimates of school effects might have "valued" two schools with very different SES composition as equally *effective* when in fact there was a wide *proficiency* gap between them. To put it bluntly, even if proven more robust, classifying according to models that give special treatment to school effects might have conveyed the (false) impression that students in poor but effective schools achieve as much reading, math, and sciences skills as those in rich but ineffective schools, when in reality the opposite held true.

Thus, what the majority of experts in charge of quality assurance tacitly established was that the optimal calibration of the SOM was one that expressed a fine-tuned balance between scientific conventions about the technical aspects of quality valuation and socially accepted

conventions that the education that really “qualifies” as “valuable” is, ultimately, the kind of education privileged kids get in the schools they attend. What, from the prevailing standpoint, seemed to make the ordinalization “legitimate” was not (only) that it conformed to the best standards of scientific expertise, but also that it better approximated commonsense expectations about the actually existing social distribution of learning, sanctioning schools’ academic prestige accordingly. The quest for “fairness”, in the end, concerned much more this moral commitment to matching each school’s statistically predicted value with its socially expected value, than honoring the professional maxim of delivering the most scientifically valid measure of school quality.

Still unresolved is the puzzle of why the minority of progressive experts acquiesced to a version of school accountability that fell short of its expectations for fair responsabilization. Analyzing experts’ reasoning through the lens of Abend’s [2014] theory of nested orders of morality, I have shown that a common moral background facilitated accommodation. Despite their moral qualms about choosing a school valuation method that might not fit a particular conception of fair responsabilization, policymakers’ recourse to the quasi-experimental methods of applied microeconomics—the empirical estimation of an unbiased parameter thought to statistically identify school quality—proved critical to framing both the terms of and solutions to this controversy [Breslau 2013; Panhans and Singleton 2017]. This moral background, I contend, provided “the criteria for morality’s or moral considerations’ being relevant in a situation in the first place” [Abend 2014, 30], in this case, whether in the evaluation of the SOM’s fairness morality was at all applicable. In this sense, experts shared a faith in scientific procedures (e.g., econometrics) as appropriate tools for evaluating fairness, a moral category. They differed in their preferred empirical method of performing an index of school quality. Whether an empirical method was necessary to base judgments on fairness was not under question.

More importantly, the shared scientism that facilitated experts’ deliberations also implied that the definition of the “school valuation problem” as an explicitly moral one was difficult to articulate. Although they repeatedly invoked vague notions of goodness and badness, rightness and wrongness, correctness and incorrectness in order to attribute different “value” to different objects (schools vs. students), they did not frame their acts of value attribution as moral practices. Had they proceeded otherwise, they would have likely compromised their authority as experts endowed with a sense of scientific objectivity and, therefore,

moral impartiality. Thus, what Abend calls “moral grounding”—the substantive reasons that justify peoples’ moral judgments—assumed in the case of the SOM a paradoxical form. As guardians of the rules of mechanical objectivity [Porter 1996], experts were expected to avoid engaging in moral argument. Their moral background therefore dictated that there must be no ground for moral grounding. When grounding judgment (as in the case of SES adjustments or proficiency thresholds) required giving moral reasons, they were conveniently camouflaged as technical decisions. Certainly, such technical decisions were openly deemed good or bad insofar as they measured the quality of schools fairly or not. But neither the interpretation of the legal mandate for fairness nor the way of technically handling it that finally prevailed were overtly framed as the object of a pluralistic, rational moral evaluation [Albury 1983]. Indexed to an allegedly indisputable notion of quality—one that safeguarded proficiency benchmarking against an unrestricted application of the SES adjustment mandate—the SOM was rendered both technically and morally infallible.

Experts’ moral background, finally, comprised a tacit agreement that only consequentialist arguments count as valid “technical” claims. This tacit agreement drove the importance of simulations of the distributive consequences associated with different calibrations of the SOM in adjudicating fair responsabilization. Consequentialism, moreover, presupposed a realist social ontology: the convention that “objects preexist—at least in theory—the work of identification, definition, and delimitation” [Desrosières 2002: 307]. The statement, decreed by executive order, that what remains unaccounted for by the regression model—the purely statistical artifact of the equation’s error term—is an indicator of school effectiveness, bears witness to this realism. Ultimately, the SOM’s performative power as policy device laid in an *econometric* style of reasoning and doing that construed “education quality” as something that exists independently of the methods for its measurement, in the form of a single, quantifiable trait unevenly distributed across a population of schools that could (and therefore must) be hierarchically sorted accordingly [Espeland and Sauder 2007]. More consequentially, the variance-partitioning approach espoused to account for this trait further rationalized and reified school practices as an “environmental” effect that, albeit unobservable (hence its definition as a residual), could nevertheless be statistically delimited, calculated, and thus treated independently of “inherited” social endowments, thereby refurbishing a reversed version of the traditional nature-vs.-nurture debate [Tabery 2014; Carson 2007] whereby student-SES came to be seen as a sort of school’s “genetic makeup.”

Again, the only thing experts did not share was a common definition of fairness itself, a moral disagreement of first order. Paradoxically, progressives leaned towards a classic meritocratic argument (with the only caveat that merit in this case was assessed at the organizational rather than the individual level): unfairness meant disregarding considerations of schools' merit in the attainment of student proficiency. Valuing school efforts mattered, especially considering the history of discrimination and abandonment of failing public schools after decades of state-sponsored privatization. Thus, merit-based valuation was necessary to compensate for a history of unfair treatment.

Market-minded experts' standpoint was, in contrast, radically consequentialist: whether schools were strictly valued for their merit mattered less than whether standards of proficiency were universally enforced in all cases. They equated unfairness with the adoption of double standards in school valuation—an unacceptable signal of market distortion. Because what was fair to poor schools was unfair to their students—the ultimate subjects of the right to quality education—the “valuation problem” must be resolved on behalf of the latter.

Conclusions

By examining how normative preferences regarding the marketization of social policy (e.g., education) percolate into market experts' work of market evaluation and repair, this research confirms what critical theory has long argued at least since Habermas [1971], namely, that technocrats' scientific judgements are neither monolithic nor impartial. As a policy device, the SOM was devised to help regulators “see and make decisions about the world in economic ways” [Hirschman and Berman 2014: 782], making market agents governable at a distance [Rose and Miller 2010]. Governing the market became technically possible inasmuch as experts rendered the SOM capable of performing—in the sense of simultaneously declaring and rendering real [Callon 2007, 1998, MacKenzie and Millo 2003, Muniesa, Millo, and Callon 2007, Muniesa 2014]—“objective” categorizations of quality. This capacity to produce “objective” quality valuations, as opposed to “subjective” assessments of school performance based on value judgements, was what ultimately imbued the SOM with “value” as a market regulation device—a capacity that was in turn grounded in a set of quantification rules and procedures of “mechanical objectivity” [Porter 1996].

Yet the scientific/technical character of the SOM as a valuation device did not prevent it from being regarded as value-neutral. Rather, since the onset, experts reflexively “saw” the SOM as embodying an “intensely moralized” technopolitical project aimed at the creation of within-market “classification situations”: technologically generated market positions that are highly consequential in terms of the classified’s opportunities and experience [Fourcade and Healy 2013, 560, 2017, 10]. As this study reveals, although policymakers explicitly mandated the SOM to perform presumably “fair” quality valuations, the clash of incommensurable conceptions of fairness rendered this task quite intractable. The quest for mechanical objectivity could not carry on without expert judgments and conventions informing rule-making. Just as objective methods of quantification appeared necessary to rule over subjects equally entitled to diverse moral judgments [Porter 1996], so were moral judgements necessary to rule over objective methods of quantification. In arbitrating competing definitions of fairness, ideologically grounded appeals to morality were experts’ last resort.

The imperative of objectivity and impartiality, therefore, did not preclude the technocrats from weaponizing their calculations to suit their moral stances on the use of market devices to address collective concerns [Frankel, Ossandón, and Pallesen 2019]. At stake in the differing calibrations of the instrument (OLS vs. HLM) was how much they valued the resulting ordinalization as a device that could serve the cause of market education. Political calculations entangled with econometric ones. Cognizant of the field-level effects of their ordinalization instrument, they vied for making it more or less friendly for families (the consumers), and more or less “nudging” for schools (the providers). Thus, belief in whether the policy field was to be organized more or less in a market or bureaucratic fashion framed the process of model evaluation.

And yet, the fact that market designers still managed to bridge the ideological gap and proceeded to implement the SOM even as the political question of the market organization of the policy field could not be dodged is theoretically puzzling. I tackled this puzzle aided by Abend’s conceptualization of nested orders of morality. In detailing how an “econometric” moral background, prevalent among the social-science experts, was pivotal in steering the adjudication of competing moral choices, this research sheds light on the inner workings of neoliberal resilience. Scholars have argued that the persisting tendency to frame policy problems in terms of efficiency—and thus give preference to markets as policy solutions—cannot be understood without taking heed

of economists' growing technocratic leverage across governments and international organizations [Berman 2021; Fourcade 2006; Mudge 2015]. Proof of the superiority [Fourcade; Ollion and Algan 2015] of economic styles of reasoning is the widespread institutionalization of policy devices that, like the SOM, make the world governable in economic ways [Hirschman and Berman 2014]. This paper underscored the *procedural* aspects of this style of reasoning. In a way, the quest for factoring class inequality into school quality assessments in Chile fits Reverdy and Breslau's [2019: 215] characterization of the process of market entrenchment as one in which market experts continuously strive "to reframe a political question of distribution and justice as an economic question of the functioning of a market." Thus, neoliberalism owes its resilience not to the technocrats' capacity to immunize themselves from external demands for decommodification, but to their perseverant work of translating claims against the market's political failure into claims of market failure in the economic sense [Reverdy and Breslau 2019, 215]. Econometric reasoning—conceptualizing competing ends in terms of calculable tradeoffs, simulating such tradeoffs through modeling and adjusting, and making decisions based on the simulated distributive consequences—serves as the lingua franca for this translation. It not only helps technocrats (whether economists or other social science experts) frame policy problems and solutions in economic terms; it also equips them with the cognitive and procedural means for dealing with conflicting policy frames.

BIBLIOGRAPHY

- ABEND Gabriel, 2014. *The Moral Background: An Inquiry Into The History Of Business Ethics* (Princeton University Press).
- ALBURY W.R., 1983. *The Politics of Objectivity* (Deakin University Press).
- AMABLE Bruno, 2011. "Morals and politics in the ideology of neo-liberalism," *Socio-Economic Review*, 9 (1): 3–30.
- AMERICAN-EDUCATIONAL-RESEARCH-ASSOCIATION, 2015. "AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs," *Educational Researcher*, 44 (8): 448–452.
- AMERICAN-STATISTICAL-ASSOCIATION, 2014. *ASA statement on using value added models for educational assessment*.
- BECKERT Jens, 2009. "The social order of markets," *Theory and society*, 38 (3): 245–269.
- BENVENISTE Luis, 2002. "The political structuration of assessment: Negotiating state power and legitimacy," *Comparative Education Review*, 46 (1): 89–118.
- BERGBAUER ANNIKA B, ERIC A HANUSHEK and Ludger WOESSMANN, 2018. "Testing," *National Bureau of Economic Research, Working Paper*, 24836 (July 2018).
- BERMAN ELIZABETH Popp, 2021. *Thinking Like an Economist: How economics became the language of U.S. public policy* (Princeton, NJ, Princeton University Press).
- BOLTANSKI Luc and Arnaud ESQUERRE 2016. "The Economic Life of Things. Commodities,

- Collectibles, Assets," *New Left Review*, (98): 31–56.
- BOLTANSKI Luc and Laurent THÉVENOT, 2006. *On justification: Economies of worth* (Princeton, Princeton University Press).
- BRESLAU Daniel, 2013. "Designing a market-like entity: Economics in the politics of market formation," *Social Studies of Science*, 43 (6): 829–851.
- CALLON Michel, 1998. *The Laws of the markets*, vol. 6 (Oxford, Blackwell).
- , 2007. "What does it mean to say that economics is performative?" in Donald MacKenzie, Fabian Muniesa and Lucia Siu, eds, *Do economists make markets?: on the performativity of economics* (Princeton, Princeton University Press: 311–357).
- CARSON John, 2007. *The Measure of merit: Talents, intelligence, and inequality in the French and American republics, 1750–1940* (Princeton, Princeton University Press).
- CORBALÁN PÖSSEL Francisca and Patricio Corbalán CARRERA, 2012. "El rol del think-tank. Libertad y desarrollo en la consolidación de la educación neoliberal en Chile," *Profesorado*, 16 (3): 169–190.
- DAVIES W., 2016. *The Limits of Neoliberalism: Authority, Sovereignty and the Logic of Competition* (London, SAGE Publications).
- DESROSIERES Alain, [1993] 2002. *The Politics of large numbers: A history of statistical reasoning* (Cambridge/London, Harvard University Press).
- DONOSO Sofia, 2013. "Dynamics of Change in Chile: Explaining the Emergence of the 2006 Pinguino Movement," *Journal of Latin American Studies*, 45 (01): 1–29 [doi: 10.1017/S0022216X12001228].
- ESPELAND W. and Michael SAUDER, 2007. "Rankings and reactivity: How Public Measures Recreate Social Worlds," *American journal of sociology*, 113 (1): 1–40.
- ESPELAND W. and Mitchell L. STEVENS, 1998. "Commensuration as a social process," *Annual review of sociology*, 24 (1): 313–343.
- FALABELLA Alejandra, 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.
- FOURCADE M., 2006. "The Construction of a Global Profession: The Transnationalization of Economics," *American journal of sociology*, 112 (1): 145–194.
- , 2011. "Cents and sensibility: Economic valuation and the nature of 'nature'," *American journal of sociology*, 116 (6): 1721–1777.
- , 2016. "Ordinalization: Lewis A. Coser Memorial Award for Theoretical Agenda Setting 2014," *Sociological Theory*, 34 (3): 175–195.
- FOURCADE M. and K. HEALY, 2013. "Classification situations: Life-chances in the neoliberal era," *Accounting, Organizations and Society*, 38 (8): 559–572.
- FOURCADE M., and K. HEALY, 2017. "Seeing like a market," *Socio-Economic Review*, 15 (1): 9–29.
- FOURCADE Marion, Etienne OLLION and Yann ALGAN, 2015. "The superiority of economists," *Journal of economic perspectives*, 29 (1): 89–114.
- FRANKEL Christian, José OSSANDÓN and Trine PALLESEN, 2019. "The organization of markets for collective concerns and their failures," *Economy and Society*, 48 (2): 153–174.
- GORUR Radhika, 2016. "Seeing like PISA: A cautionary tale about the performativity of international assessments," *European Educational Research Journal*, 15 (5): 598–616.
- GREK Sotiria, 2009. "Governing by numbers: The PISA 'effect' in Europe," *Journal of education policy*, 24 (1): 23–37.
- GRIFFEN Zachary, 2020. "The 'production' of education: the turn from equity to efficiency in US federal education policy," *Journal of Education Policy*: 1–19.
- HABERMAS, Jürgen, 1971. *Toward a rational society: Student protest, science, and politics*, vol. 404 (Boston, Beacon Press).
- HACKING Ian, 2002. "Inaugural lecture: Chair of philosophy and history of scientific concepts at the Collège de France, 16 January 2001," *Economy and society*, 31 (1): 1–14.
- HANUSHEK ERIC A., 1986. "The economics of schooling: Production and efficiency in public schools," *Journal of economic literature*, 24 (3): 1141–1177.
- , 2009. "Building on no child left behind," *Science*, 326 (5954): 802–803.
- HARRIS, Douglas N., 2011. *Value-Added Measures in Education: What Every Educator Needs to Know*: ERIC.
- HASAN Sharique and Anuj KUMAR, 2019. "Digitization and divergence: Online school ratings and segregation in America," *Social Science Research Network* [Available at SSRN: <https://ssrn.com/abstract=3265316> or <http://dx.doi.org/10.2139/ssrn.3265316>].
- HIRSCHMAN ALBERT O., 1970. *Exit, voice, and loyalty: responses to decline in firms*,

- organizations, and states*, vol. 25 (Harvard, Harvard university press).
- , 1991. *The rhetoric of reaction* (Cambridge, MA, Harvard University Press).
- HIRSCHMAN Daniel and Elizabeth Popp BERMAN, 2014. "Do economists make policies? On the political effects of economics," *Socio-Economic Review*, 12 (4): 779–811.
- KARPIK Lucien, 2010. *Valuing the unique: The economics of singularities* (Princeton, Princeton University Press).
- LARROULET C. and P. MONTT, 2010. "Políticas educativas de largo plazo y acuerdo amplio en educación: el caso chileno," in Sergio Martinic and Gregory Elacqua, eds, *Fin de Ciclo: Cambios en la Gobernanza del Sistema Educativo* (Santiago de Chile, UNESCO).
- LATOUR Bruno, 1987. *Science in action: How to follow scientists and engineers through society* (Cambridge Mass., Harvard University Press).
- LIVNE Roi, Daniel HIRSCHMAN, Margaret R. SOMERS and Elizabeth Popp BERMAN, 2018. "In search of social backgrounds. On Gabriel Abend's, The Moral Background, An Inquiry into the History of Business Ethics, Princeton, NJ, Princeton University Press, 2014," *Socio-Economic Review*, 16 (3): 637–655.
- MACKENZIE Donald and Yuval MILLO, 2003. "Constructing a market, performing theory: The historical sociology of a financial derivatives exchange," *American journal of sociology*, 109 (1): 107–145.
- MANGEZ Eric and Mathieu HILGERS, 2012. "The field of knowledge and the policy field in education: PISA and the production of knowledge for policy," *European educational research journal*, 11 (2): 189–205.
- MAXWELL JOSEPH A., 2012. *Qualitative research design: An interactive approach: An interactive approach* (Thousand Oaks Cal., Sage).
- MEHTA Jal, 2013. *The Allure of Order: High Hopes, Dashed Expectations, and the Troubled Quest to Remake American Schooling* (New York, Oxford University Press).
- MUDGE STEPHANIE Lee, 2008. "What is neoliberalism?," *Socio-Economic Review*, 6 (4): 703–731 [doi: 10.1093/ser/mwn016].
- , 2011. "What's Left of Leftism?: Neoliberal Politics in Western Party Systems, 1945–2004," *Social Science History*, 35 (3): 337–380.
- , 2015. "Explaining Political Tunnel Vision: Politics and Economics in Crisis-Ridden Europe, Then and Now," *European Journal of Sociology*, 56 (01): 63–91.
- MUNIESA Fabian, 2014. *The provoked economy: Economic reality and the performative turn* (London/New York, Routledge).
- MUNIESA Fabian, Yuval MILLO and Michel CALLON, 2007. "An introduction to market devices," *The Sociological Review*, 55:1–12 [doi: 10.1111/j.1467-954X.2007.00727.x].
- MUÑOZ-CHEREAU B. and S.M. THOMAS, 2016. "Educational effectiveness in Chilean secondary education: comparing different 'value added' approaches to evaluate schools," *Assessment in Education: Principles, Policy & Practice*, 23 (1): 26–52.
- NIK-KHAH Edward and Philip MIROWSKI, 2019. "On going the market one better: economic market design and the contradictions of building markets for public purposes," *Economy and Society*, 48 (2): 268–294.
- OECD, 2013. *Synergies for better learning: An international perspective on evaluation and assessment*.
- OECD, 2016. *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools* (Paris, OECD).
- OSSANDÓN José and Sebastián URETA, 2019. "Problematising markets: market failures and the government of collective concerns," *Economy and Society*, 48 (2): 175–196.
- PANHANS MATTHEW T. and John D. SINGLETON, 2017. "The Empirical Economist's Toolkit: From Models to Methods," *History of Political Economy*, 49: 127–157.
- PODOLNY JOEL M., 1993. "A status-based model of market competition," *American journal of sociology*, 98 (4): 829–872.
- PORTER THEODORE M., 1996. *Trust in numbers: The pursuit of objectivity in science and public life* (Princeton, Princeton University Press).
- RAUDENBUSH STEPHEN W. and J. Douglas WILLMS, 1995. "The estimation of school effects," *Journal of educational and behavioral statistics*, 20 (4): 307–335.
- REVERDY Thomas and Daniel BRESLAU, 2019. "Making an exception: market design and the politics of re-regulation in the French electricity sector," *Economy and Society*, 48 (2): 197–220.
- ROSE N. and P. MILLER, 2010. "Political power beyond the state: Problematics of government," *The British Journal of Sociology*, 61: 271–303.

THE MORAL LIFE OF ECONOMETRIC EQUATIONS

- SAN MARTÍN Ernesto and Alejandro CARRASCO, 2012. "Clasificación de escuelas en la nueva institucionalidad educativa: contribución de modelos de valor agregado para una responsabilización justa," *Temas de la Agenda Pública*, 7 (53): 1-16.
- SOMERS MARGARET R. and Fred BLOCK, 2005. "From poverty to perversity: Ideas, markets, and institutions over 200 years of welfare debate," *American Sociological Review*, 70 (2): 260-287.
- TABERY James, 2014. *Beyond versus: The struggle to understand the interaction of nature and nurture* (Cambridge Mass./London, MIT Press).
- TEIXEIRA PEDRO Nuno, 2000. "A Portrait of the Economics of Education, 1960-1997," *History of Political Economy*, 32 (Suppl. 1): 257-288.
- TODD PETRA E. and Kenneth I. WOLPIN, 2003. "On the specification and estimation of the production function for cognitive achievement," *The Economic Journal*, 113 (485): F3-F33.

Résumé

La recherche sociologique récente sur la conception du marché est mal outillée pour comprendre les aspects normatifs et politiques des pratiques des experts en relation avec les conflits politiques sur la marchandisation des droits sociaux. Je développe une approche originale de l'utilisation politisée des dispositifs de marché pour répondre aux préoccupations collectives dans un domaine politique non économique: l'éducation. Lors de la conception d'un important système de responsabilisation scolaire, les décideurs politiques au Chili ont été confrontés à une énigme morale: les écoles devraient-elles être évaluées en fonction de la compétence absolue de leurs élèves ou en fonction de l'efficacité relative de l'école? Les experts progressistes et conservateurs chargés de résoudre ce dilemme ont poussé à utiliser le modèle statistique (régressions linéaires contre modèles multiniveaux) donnant des classements qui correspondent à leurs préférences morales. Avec des analyses qualitatives de l'application concrète des méthodes quantitatives par les experts, ainsi que des interprétations par les experts des conséquences performatives de ces méthodes, je mobilise le concept discuté de « contexte moral » pour mettre en évidence les conditions qui permettent de subsumer la dissidence idéologique sous des formes consensuelles de prise de décision.

Mots-clés: Accountability; Dispositifs de marché; Experts; Moralité; Néolibéralisme.

Zusammenfassung

Die neuere soziologische Forschung zum Marktdesign ist kaum in der Lage, die normativen und politischen Aspekte der Expertenpraktiken bezüglich der politischen Konflikte in Sachen Vermarktung sozialer Rechte zu begreifen. Ich entwickle einen originellen Ansatz für die politisierte Nutzung von Marktinstrumenten, um kollektive Anliegen in einem nicht-ökonomischen Politikbereich zu untersuchen: Bildung. Bei der Entwicklung eines umfassenden Schulkontrollsystems standen die politischen Entscheidungsträger Chiles vor einem moralischen Dilemma: Sollten Schulen nach der absoluten Kompetenz ihrer Schüler oder nach der relativen Effektivität der Schule bewertet werden? Progressive und konservative Experten, die mit der Lösung dieses Dilemmas beauftragt wurden, drängten darauf, statistische Modellierungen (lineare Regressionen versus mehrstufige Modelle) zu verwenden, deren Rangfolgen ihren moralischen Präferenzen entsprachen. Dank qualitativer Analysen der konkreten Anwendung quantitativer Methoden durch Experten sowie deren Interpretationen der performativen Konsequenzen dieser Methoden mobilisiere ich das diskutierte Konzept des „moralischen Kontextes“, um die Bedingungen aufzuzeigen, die eine Überführung ideologischer Meinungsverschiedenheiten in eine konsensuelle Form der Entscheidungsfindung erlauben.

Schlüsselwörter: Rechenschaftspflicht; Marktinstrumente; Experten; Moral; Neoliberalismus.