How are the 'losers' of the school accountability system constructed in Chile, the USA and England?

Keywords

High-stakes testing, school accountability, performance-based accountability, failing schools, disadvantaged communities

Abstract

Performance-based accountability systems that rank schools based on their effectiveness produce 'winners' and 'losers'. Substantial evidence has pointed to the (side)effects of these classifications, particularly in the most disadvantaged communities. Whilst previous studies have compared schools under different effectiveness categories within and between countries, this qualitative study takes a cross-case comparison approach to analyse education policies, grey literature and previous research evidence to explore the mechanisms that construct 'failing' schools in three notable high-stakes accountability systems worldwide: Chile, the USA and England. After describing (1) the identification and classification of 'failing' schools; (2) the instruments used to justify these classifications; and (3) who make these judgements, we conclude that the construction of 'failing' schools serves the competition and differentiation required for maintaining neoliberal hierarchical and unequal market-oriented regimes. Instead of disciplining only 'failing' schools, these labels also provide a caution for the whole school system.

Introduction

During the last decade performance-based/high-stakes accountability systems have proliferated around the world (OECD 2013; Falabella 2020). Most countries across Europe and many others worldwide with growing decentralised market-oriented educational systems have developed their own frameworks intending to promote school change, improvement and

innovation specifically (Ehren and Shackleton 2016), and support evidence-based governance and accountability more generally (Altrichter and Kemethofer 2015). Despite the pervasiveness of these systems, results of their effects on school outcomes are mixed. Studies of various inspection categories have frequently reported no effects or improvements in compulsory core subjects measured by standardised tests for sub-groups of students (Shaw et al. 2003; Rosenthal 2004; Luginbuhl, Webbink, and de Wolf 2009; Ehren and Shackleton 2016). Moreover, research has increasingly suggested that these systems 'come at a cost, as different side effects may occur' (Penninckx et al. 2016, 335).

Whilst their scope, methods, standards, data and instruments vary greatly from one governance system to another (Faubert 2009), each system develops a different strategy when dealing with 'failing' schools. Although not all systems label schools as 'failing', each one uses deficit terminology (such as *underperforming*, *inadequate* or *very weak*) to refer to schools that score below expected standards and are positioned at the bottom of the hierarchy. Moreover, the overall discourse around these schools blames them for their own failure. Thus, by implementing accountability systems that judge and rank schools on the basis of their effectiveness, *winners* and *losers* are produced (Schagen and Schagen 2003; Greany and Higham 2018). The *losers* of the accountability game 'get represented in ways that mystify power relations and often create individuals responsible for their 'failures', drawing attention away from the structures that create unequal outcomes' (Bacchi 2000, 46). Meanwhile, the structural inequities of the system and the working conditions within these institutions remain invisible (Falabella 2014).

Whilst the majority of studies conducted in the last few decades in the field of school effectiveness and improvement focus on effective schools or the *what works* agenda (Munter, Cobb and Shekell 2016), a comparatively rare and limited number has addressed schools positioned at the bottom of the pile (Chapman and Harris 2004). The need to describe how different accountability systems construct the weakest schools is necessary for many reasons. The first one derives from distributive and recognitional dimensions of social justice, as these systems produce unequal distribution of knowledge and stigmatisation of schools serving vulnerable students (Flórez and Rozas 2020). For instance, in England high-stakes consequences for those schools that are classified as 'failing' are negatively affecting the population that tends to be the most disadvantaged (Hutchinson 2016; Greany and Higham 2018). The second reason is conceptual, as school failure is a relative term defined in relation

to an expected norm (Barber 2002). If the threshold changes, the nature and magnitude of the problem also changes. The third reason is policy-oriented: because school failure appears as an impossible issue to solve, by comparing different systems, the problem can be redefined. Therefore this paper is timely, as previous studies have compared schools under different effectiveness categories within a given country (Penninckx et al. 2016) or between countries (Ehren et al. 2015), but no previous study has explored the way in which a similar phenomenon is created in three notable high-stakes school accountability systems (Falabella and De La Vega 2016): Chile, the USA and England. We argue that despite policy variations—resulting from each country's specific contexts and history—shared trends can be identified.

Literature review

To situate the discussion of the *losers* of the accountability game within its wider context, we will shed light on the performative shift in governance specifically, and neoliberal market-oriented reforms that nurtured its emergence generally. In this scenario, the *performing school* is configured by a market- and state-accountability model where practitioners (school leaders and teachers) need to continuously perform successfully within a predominantly competition-based framework (Gleeson and Husbands 2001; Maroy 2009; Falabella 2014). This performance-based accountability system relies on the implementation of high-stakes testing, which is then used to hold school staff accountable for their institutional outcomes and to determine potential consequences in terms of rewards and sanctions according to their attainment level. Falabella (2020, 24) explains that 'this policy approach is based on the use of quantifiable indicators, systems of comparison, and external dis/incentives on the expectation that these measures will positively encourage school staff to continuously promote educational quality for all'.

Closely aligned, neoliberal market-oriented reforms produce decontextualized and apparently objective indicators on the basis of which schools' performance is classified and compared. A particular type of data-based knowledge governance grounded in choice and competition—the two principles on which neoliberal market-oriented reforms rely—are made possible and manageable (Ozga 2016). Meanwhile, a wake-up call stressing the detrimental effects that these reforms are having on educational systems, particularly from social justice perspectives

concerned with the social goals of education, equity and social cohesion (Levin 2018), has increasingly been articulated (Falabella 2014; Flórez and Rozas 2020). Despite neoliberal market-oriented reforms driving some productive efficiencies reflected in increases in average student achievement scores (Rouse et al. 2013; Saw et al. 2017), they tend to be coupled with disastrous consequences in terms of unequal opportunities and stigmatisation of disadvantaged school communities. Over time fewer educational opportunities for the most disadvantaged students are afforded and high levels of socio-economic segregation become entrenched (OECD 2013).

Some scholars, taking a post-critical sociological framework (Falabella 2020), argue that it is misleading to interpret these effects as secondary or unintended, as they are the expected outcomes of the neoliberal policy regime (Harvey 2005; Ozga 2016; Falabella 2014, 2020; Flórez and Rozas 2020). Thus, classifying and labelling schools as 'failing' in the name of transparency turns into a powerful social control practice oriented to discipline disadvantaged groups (Ozga 2016). 'The key point is that the model is based on a competitive rationale, which accepts, and moreover, requires the existence of a hierarchical unequal field with institutions that are differently positioned (…) If all schools were to attain the same benchmark, the model would not work' (Falabella 2014, 12).

The call for moving away from performance-based accountability has also been posed by researchers documenting the handicaps derived from the use of big data in general, and standardised tests in particular. By revealing spurious past correlations between educational variables, standardised tests neglect *small data*, understood as the details that make the difference in schools: how good quality teaching leads to better learning. In this view, school-based and formative assessments are much more likely to improve the quality of education because they capture in their narratives the richness and complexity of the details and relationships that matter at school (Sahlberg and Hasak 2017).

During the last decade attention has also been paid to the ways in which accountability systems affect schools differently within a country. That is, the impact an accountability system has on a given school partly depends on the overall judgement of its effectiveness (Penninckx et al. 2014) or its place in the vertical hierarchy (Falabella 2014). There is evidence that 'differential degrees of "accountability pressure"' on school leaders (Altrichter and Kemethofer 2015, 32) and teachers (Penninckx et al. 2016) are a result of inspections and assessments. Schools deemed to be 'failing' usually receive greater levels of pressure

(Falabella 2014) and embark on differing patterns of improvement of student performance after inspection, depending on whether the school is judged to be weak, average or strong (Matthews and Sammons 2005). 'Failing' schools tend to narrow the curriculum, teach to the test and avoid innovation, which in turn worsen teaching and learning (Flórez and Rozas 2020).

Materials and Methods

Study Design

We explored the mechanisms that construct 'failing' schools in three jurisdictions that feature high-stakes accountability systems (Falabella and De La Vega 2016):Chile, the USA and England. As Scotland, Wales and Northern Ireland have their own accountability systems that are not necessarily considered high-stakes, we focus on England rather than the UK. Similarly, given the diversity among state systems across the USA, we draw on Virginia as a relatively representative example of state system. Our qualitative exploratory study was framed by the following research questions:

1. Who makes the decision about 'failing' schools and at what level are decision-makers located?
2. What is the process of identifying 'failing' schools?
3. What instruments are used to identify 'failing' schools?

We adopted a cross-case comparison design to develop thick descriptions of each case and subsequently compare them to identify cross-case patterns (Miles, Huberman and Saldaña 2014). A cross-case comparison design affords a deep investigation of a single phenomenon (i.e., how 'failing' schools are constructed by accountability systems) in two or more cases, retaining holistic and meaningful characteristics of each studied case while also leading to tentative generalisations through their comparison (Rihoux and Bojana 2009).

Data sources and analysis

This comparative qualitative study analyses secondary data. More specifically, we conducted desk-based research oriented to examine source documents, such as external evaluation and

inspection guidelines, frameworks, handbooks, policy documents and legislation, and relevant academic literature on each of the three studied jurisdictions (examined documents, reports and papers are marked with * in the Reference section). To explore the ways in which each accountability system constructed 'failing' schools, we developed a two-step analytical process: within-case analysis and cross-case comparisons (Miles, Huberman and Saldaña 2014). Regarding within-case analysis, authors answered each research question focusing on the jurisdiction where s/he had the most expertise by selecting, reading and coding documents in order to identify themes referring to accountability in underperforming schools that allow the production of a focused, thick case description.

Second, the cross-case comparisons examined the data across the three jurisdictions in order to reveal commonalities and differences (Brooks Hall 2017). For that purpose, extensive discussions were maintained between the authors through two face-to-face meetings between the first and second author, fourteen video calls between all authors, and many email exchanges to discuss the findings.

Results

Within-case analysis

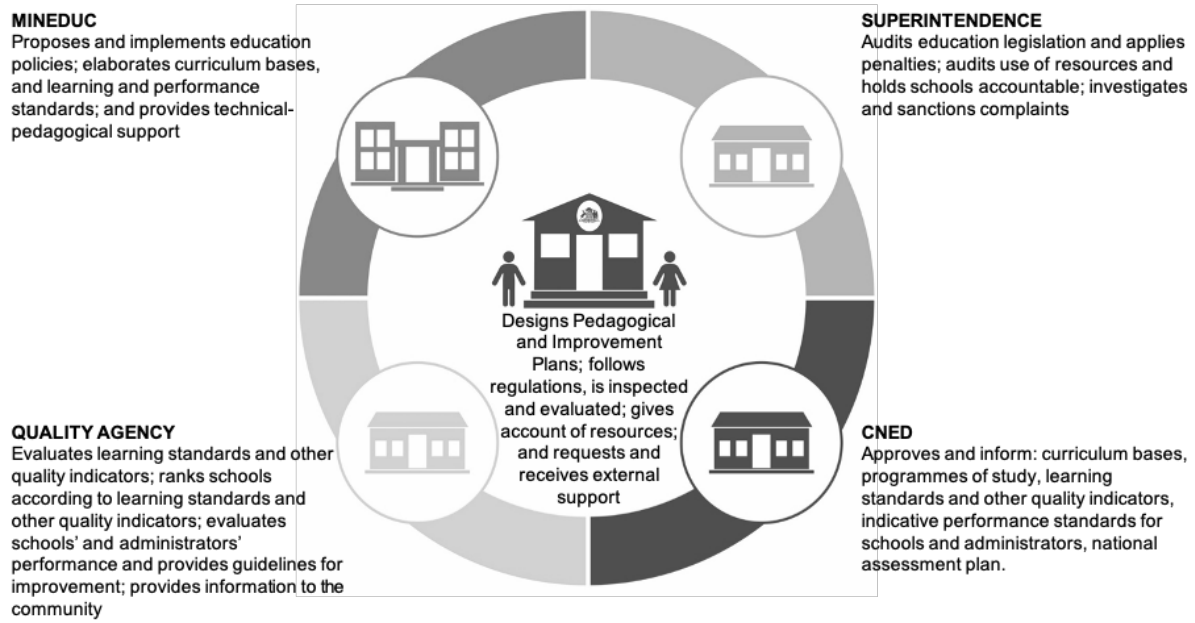
'Failing' schools in Chile

Chile has established a decentralised education system configured as a market grounded on parental choice and privatisation, where funding is provided by the state to municipal and private school administrators through a voucher based on students' attendance (Bellei and Vanni 2015). Additionally, the Chilean system introduced test-based/high-stakes accountability to address quality assurance and school improvement (Ahumada, Montecinos and González 2012). Since the 1980s, schools' performance is assessed by the central government through SIMCE, a census-based standardised test that annually appraises students' learning in all types of schools according to national curriculum standards (Meckes and Carrasco 2010).

The core of current school improvement policy is contained in two legal documents. First, law Nº 20,248 from 2008 sets a Preferential School Subsidy (SEP) or adjusted voucher for

students from disadvantaged families attending state-funded schools (MINEDUC 2008a). As a condition to receive this additional funding, schools must design a four-year School Improvement Plan and are accountable for disadvantaged students' performance in SIMCE (MINEDUC 2008b). Second, law Nº 20,529 from 2011 sets up the National System of Quality Assurance of Education (SAC), which introduced two new government agencies: The Superintendence of Education, which audits schools' and their administrators' compliance with legislation; and the Education Quality Agency (ACE), which evaluates schools' performance and carries out inspection visits targeting low-performing schools.

The structure of the quality assurance system (SAC) is depicted in Figure 1.

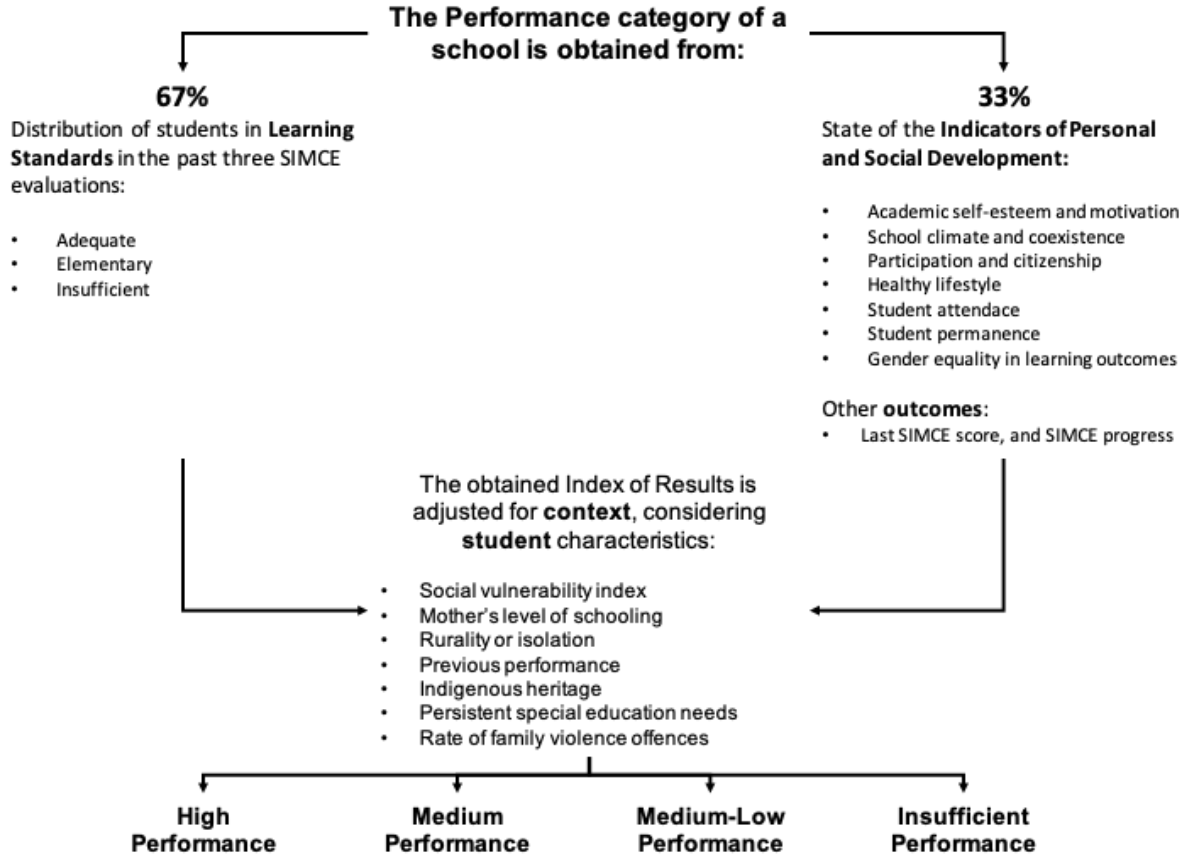


Who makes the decision about 'failing' schools in Chile and at what level are decision-makers located? Within the institutional arrangement of SAC, the mission of ACE is to evaluate, guide and inform the educational system to focus on improving the quality and equity of educational opportunities (ACE 2017). In terms of evaluation, ACE manages a set of national assessment instruments of academic and non-academic educational outcomes, in addition to coordinating the participation of Chile in international assessments such as PISA (Programme for International Student Assessment).. Regarding guidance, ACE develops inspection visits to provide feedback to low-performing schools about their teaching and management processes. Finally, in terms of information, ACE promotes schools' use of assessment data to inform their improvement processes, as well as informing parents and guardians about the attainment of students in a given school.

ACE is a functionally decentralised and autonomous state institution, endowed with its own assets, acting in coordination with the Ministry of Education (MINEDUC). It is headed by an executive secretary and a specially appointed council that oversees its strategic planning. To achieve its mission, ACE is internally organised into five divisions (assessment, guidance, studies, information, and administration), and territorially into five macrozones (north, centre-north, centre-south, south and austral) in order to coordinate its activities with schools across the country.

What is the process of identifying 'failing' schools in Chile? All types of schools are sorted into four Performance Categories: *High* (students perform above expected), *Medium* (students perform as expected), *Medium-Low* (students perform below expectations), and *Insufficient* (students perform significantly below expected) (ACE 2017). The Performance Category is the result of the yearly evaluation of schools' performance, allowing ACE and MINEDUC to determine possible sanctions in addition to identifying the level of support and guidance for each school. The Performance Category is assigned according to an Index of Results (see figure 2), which is based on students' learning (67%) and other academic and non-academic outcomes (33%) and is then adjusted according to students' and schools' social characteristics (socioeconomic status, rurality). Based on this adjusted index, schools are classified into the corresponding Performance Category.

Figure 2: Classification of Schools into Performance Categories



According to data published by ACE, of the 5,675 primary and 2,837 secondary schools evaluated in 2019[1], 343 (6%) and 169 (6%) respectively were assigned to the *Insufficient* category. The number of schools in this category has been decreasing since the first evaluation of primary schools' performance in 2016 (633, 11%) and of secondary schools in 2017 (338, 12%). However, 120 primary (2%) and 107 secondary (3%) schools have remained in the *Insufficient* performance category.

This system of classification has high-stakes consequences as it determines rewards and sanctions related with the degree of autonomy, intervention and support for schools. Schools classified with *Insufficient* performance are subject to external intervention programmes from MINEDUC delivered by their supervisors in key curriculum areas (e.g., Reading and Mathematics), and Evaluation and Performance Guidance Visits by ACE evaluators who

[1] Primary schools with data for less than 30 students (27%) and without information (1%) were excluded.

assess schools' processes and provide suggestions. ACE targeted inspection visits aim to strengthen the internal capacity of schools by offering guidance for their improvement plan (ACE 2017).

What instruments are used to identify 'failing' schools in Chile? As mentioned above, students' learning and other academic and non-academic outcome data are employed by ACE to calculate an Index of Results to assign schools into Performance Categories. Students' learning data are obtained from schools' SIMCE results. The 2016-2020 national evaluation plan (MINEDUC 2015) indicates that students sit for SIMCE tests in years four, six, eight (primary) and ten (secondary) in reading comprehension, writing, mathematics, science, history, geography and social studies, and English. Instead of providing information about individual students' scores, SIMCE reports the average score of the tested grade group (Meckes and Carrasco 2010). The Index of Results takes into account SIMCE scores in the three latest evaluations, and the distribution of the students in Learning Standards in the latest measurement. Learning Standards describe what students should know and be able to do to demonstrate they meet the learning objectives stipulated in the curriculum for a given grade and subject, and comprise three levels of learning: Insufficient, Elementary and Adequate (ACE 2016a). This information represents 67% of the Index of Results for a given school.

Regarding other academic and non-academic outcomes, the Index of Results also considers data from the latest SIMCE score and the progress made in the past three years in reading and mathematics for each tested grade (ACE 2016b), in addition to a set of Indicators of Personal and Social Development (IDPS) for each school. The IDPS include aspects that go beyond the domain of academic knowledge and are measured together with the SIMCE test through self-administered questionnaires for students, teachers, principals and families (ACE 2016c). These instruments collect information about the following topics: school climate and *convivencia* (social relations among school community members: students, teachers, parents, etc.); participation and citizen education; academic self-esteem and motivation; and healthy life habits. Additionally, non-academic outcomes consider schools' student attendance, dropout rates, gender equality, and vocational schools' graduation rates. This information represents 33% of the Index of Results for a given school.

'Failing' schools in the USA

The American education system balances federal or national oversight with local state control to implement policies intended to improve the nation's lowest-performing schools (Mintrop and Trujillo 2005). Federal policies and initiatives have waxed and waned in intensity for 20 years, but throughout this time the system has remained mostly decentralised. That is, the federal government makes educational policy that state education agencies (SEAs) interpret and enact with varying degrees of fidelity, and then, even within states, local education agencies further vary in how they interpret and enact policy (Manna, 2010).

Public Law 107-110 or more commonly, the No Child Left Behind Act of 2001 (NCLB 2002), was a federally prescriptive law that set expectations for schools, especially those lagging, to improve student achievement scores by requiring states to set minimum proficiency standards in mathematics and reading (Ballou and Springer 2017). States most frequently identified 'failing' schools as those with the lowest percentage of students scoring proficient on assessments, although states had considerable autonomy in determining proficiency standards (Balfanz et al. 2007). Thus, the notion of what made a 'failing' school varied widely by state.

While NCLB was still law, the Obama administration responded to a recession in part by signing into law the American Recovery and Reinvestment Act of 2009 (Pub. Law 111-5) in which billions of U.S. dollars were designated for Race to the Top and School Improvement Grants (SIG) initiatives (Dragoset et al. 2016; 2017). Both initiatives were designed to rapidly improve, or turn around, the lowest performing schools in the country. By its conclusion, the SIG initiative alone was a 7 billion U.S. dollar policy investment (Emma 2015). The identification of SIG eligible schools was more nuanced than NCLB parameters, but still relied primarily upon student achievement outcomes. The U.S. Department of Education clearly communicated an expectation for SEAs to identify their lowest 5% of schools overall, as well as lowest 5% of secondary or high schools, although each SEA determined how the lowest 5% would be calculated (Hurlburt et al. 2011). In the initial rollout, 1,107 (7.2%) of schools nationwide were identified as SIG-eligible in Tier I, and another 1,034 high schools (6.8% of all schools) in Tier II. The funding of SIG-eligible schools was disproportionally dispensed to urban schools (52.5% funded but representative of only 26.0% of all schools), lower income students (68.4% versus 44.7%) and racial/ethnic minority students (73.5% versus 45.0%).

Who makes the decision about 'failing' schools in the USA and at what level are decision-makers located? Like NCLB before it, the Every Student Succeeds Act of 2015 (ESSA) (Public Law 114-95) is a reauthorization of the 1,965 Elementary and Secondary Education Act (ESEA). Much of the core of the law remains, including annual standardised testing and an emphasis on external assessment, but further transitions federal accountability provisions to states (Penuel, Meyer and Valladares 2016), resulting in an even more decentralised system, although the core of the education system structure remains relatively the same. Figure 3 provides a simplified overview of the system as it pertains to Comprehensive Support and Improvement (CSI) schools, which is the current nomenclature for 'failing' schools.

Figure 3: System that classifies CSI schools

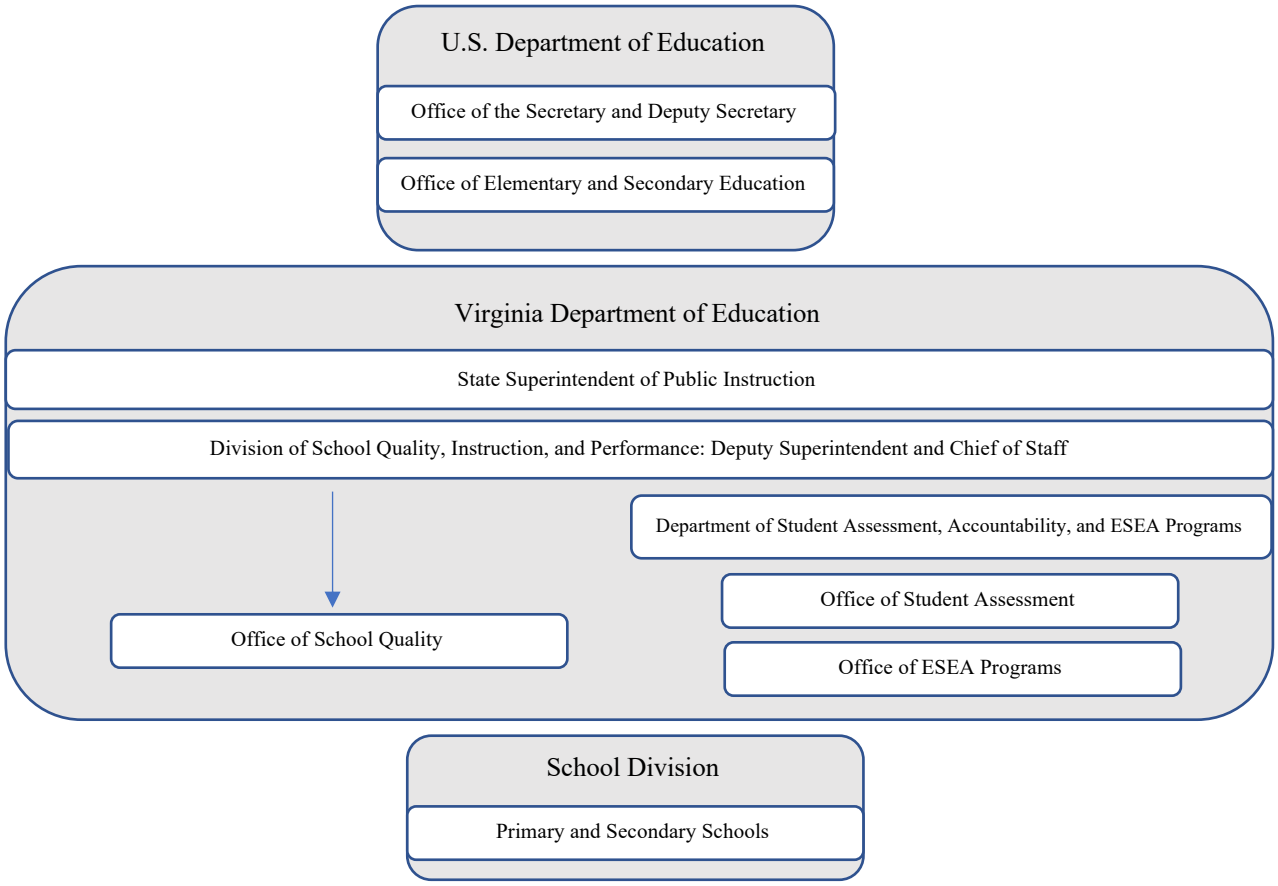


Figure 3 depicts a system in which initial policy is developed at the federal level. The enactment of the policy rests in the Office of the Secretary, with primary responsibility for identifying and improving the lowest-performing schools residing with the Deputy Secretary.

Among the many offices reporting to the Deputy Secretary, the stated mission of the Office of Elementary and Secondary Education (2020) is to 'empower States, districts, and other organizations to meet the diverse needs of every student by providing leadership, technical assistance and financial support' (para. 1) on all matters K-12 public education with a vision of 'excellence and equity for all students' (para. 2). Within it, the Office of State Support is critical to coordinating all of the various ESEA-related programmes to groups of state education agencies (United States Department of Education 2020). Each SEA is then responsible for interpreting and responding to federal policy. The state of Virginia, which is a relatively representative example of state system despite the diversity among state systems across the country (VanGronigen & Meyers 2019), has a division dedicated to issues of school quality and equity that coordinates with and responds to federal policy and reviews. This division makes many of the state-level decisions about how to measure school performance, provide and/or oversee technical assistance, and review improvement efforts (Virginia Department of Education 2018).

What is the process of identifying 'failing' schools in the USA? ESSA's transition to more state-level autonomy is a significant recent policy shift that further decentralises the identification of 'failing' schools in the USA (Rentner, Kober and Frizzell 2017). Although the federal government continues to provide similar levels of funding for the lowest-performing schools, it has considerably less oversight of the improvement processes enacted locally. Unlike NCLB, 'ESSA leaves it to states to determine how to measure school progress within certain parameters' (Rentner et al. 2017, 1). The SEA is still responsible for identifying its lowest 5% of schools and any high school with a graduation rate of less than 67%, now labelled 'Comprehensive Support and Improvement (CSI) Schools' (McGuinn 2016). The identification process must be conducted at least every three years and made publicly available (Penuel et al. 2016). The processes of school identification and public notification are typically conducted by offices of state-level accountability within SEAs.

In Virginia, for example, the state department identifies CSI schools through the following steps:

- Identify Title I schools that did not meet the interim measure of progress for English (reading), mathematics, and FGI (federal graduation index) and are in the lowest two quartiles for academic growth in English (reading) and mathematics.

- Of those schools identified in Step 1, identify schools that did not meet the interim measure for English learner progress and are in the lowest two quartiles for English learning progress.
- Of those schools identified in Step 1 through 2, identify schools that did not meet the interim measure of progress for chronic absenteeism and have an SOA (Standards of Accreditation) rating of Accredited with Conditions or Accredited Denied. Rank those schools by the rate of chronic absenteeism and identify a number equal to 5% of Title I schools for comprehensive support and improvement (Virginia Department of Education 2018).

These identification procedures and any like them in other states are approved and monitored by the U.S. Department of Education (McGuinn 2016).

What instruments are used to identify 'failing' schools in the USA? SEAs also currently have much more autonomy in how they determine which schools are the lowest-performing and if they are improving. SEAs must include at least four indicators in school ratings, three of which should be academic. Measures of student achievement, including proficiency on standardised tests in English/language arts and mathematics continue to be a central component focus, although no longer the sole one (Smith and Wright 2017). In addition to English/language arts and mathematics proficiency scores on standardized tests, states are required to incorporate additional academic indicators and non-academic indicators of 'substantial weight' (Penuel et al. 2016). In terms of other academic indicators, states must include English-language proficiency and have typically elected to include student growth in test scores as third one (McGuin 2016). This is an important distinction because traditionally low-performing schools typically have many students that can increase performance but perhaps not immediately achieve proficiency. Moreover, some states also distinguish student growth by student performance (i.e., measuring the difference in growth between higher- and lower-achieving students). In addition, the identification of high schools must incorporate graduation rates.

States must also identify at least one non-academic indicator. Frequently, states have included a measure of school climate, teacher engagement, or opportunity to learn for both students and teachers as a proxy of school quality because of their links to student outcomes such as academic achievement and social-emotional development (Penuel et al. 2016). 'States also

get to decide how to weight the individual indicators in their accountability systems though the law stipulates that academic factors have to count "much" more collectively than the others' (McGuinn 2016, 406).

'Failing' schools in England

English educational policies in the last decade have developed around the notion of the self-improving school system (Ofsted 2017). This policy encompasses academization, promotion of MATs (Multi-Academy Trusts), diminishing of local authority control of schools, and the development of school-to-school support mechanisms such as the formation of system leaders by Teaching School Alliances (Greany and Higham 2018). The greater autonomy and flexibility that the self-improving school system has is expected to foster innovation and improve educational outcomes (DfE 2016) but has also meant making the school solely responsible for its performance: what Greany and Higham (2018, 16) called 'coercive autonomy'.

Who makes the decision about 'failing' schools in England and at what level are decision-makers located? To monitor the quality of education, schools are inspected periodically by The Office for Standards in Education, Children's Services and Skills (Ofsted). Ofsted is a non-ministerial department of the UK government that reports to Parliament (Ofsted website). Since its inception in 1992, Ofsted inspection process has been revised frequently, with the latest framework introduced in September 2019. The new framework aims to better address the quality of learning and avoid unintended consequences, such as off-rolling, anxiety and stress, and curriculum narrowing (Ofsted 2019).

What is the process of identifying 'failing' schools in England? Ofsted inspects the overall effectiveness of schools by using 'all the available evidence to evaluate what it is like to be a learner in the provision' (Ofsted 2019, 8). The overall judgement focuses on four graded areas: (1) Quality of education (Ambitious intended curriculum for all students; Coherent, planned, sequenced and full intended curriculum; Support of teachers' subject knowledge; Use assessment effectively; Good and clear subject knowledge; Effective use of formative assessment and feedback; Teach to expand long term memory; Create effective environment for learning; and Implement a rigorous approach to the teaching of reading); (2) Behaviour and attitudes (Positive and respectful environment for learning; High, consistent, fair expectations on behaviour and conduct; Positive attitude to learning; High attendance); (3)

Personal development (The curriculum enhances learners' broader development; Support learners' character development; Support learners' psychical and mental health; Prepare learners for future success; Prepare learners for life in modern Britain) and (4) Leadership and management (Clear, ambitious, high-quality and inclusive school vision; Improve staff's subject, pedagogical and pedagogical content knowledge; Effective community engagement; Effective and constructive staff management; Governors hold leaders to account, and Safeguarding) (Ofsted 2019).

Table 1: Ofsted overall effectiveness judgements of primary and secondary schools over the last decade

	2016/17	2015/16	2014/15	2013/14	2012/13	2011/12	2010/11	2009/10	2008/09
Primary schools	4,120	2,468	3,655	4,823	5,847	4,636	4,250	5,037	5,323
Outstanding	19%	19%	18%	17%	17%	18%	8%	15%	16%
Good	72%	71%	67%	64%	61%	51%	47%	49%	52%
RI	8%	9%	14%	16%	19%	29%	40%	30%	29%
Inadequate	1%	1%	1%	2%	2%	2%	5%	6%	3%
Secondary schools	900	666	870	1,048	1,334	926	894	1,169	1,071
Outstanding	23%	22%	21%	21%	23%	26%	14%	19%	22%
Good	56%	56%	53%	49%	48%	40%	38%	41%	41%
RI	15%	17%	21%	23%	24%	30%	40%	31%	31%
Inadequate	6%	5%	5%	6%	5%	3%	8%	8%	6%

Source: Created by the authors from figures reported by Annual Ofsted Inspection Reports

Table 1 shows the distribution of schools according to Ofsted judgements. In 2016/17, 1% of primary and 6% of secondary schools were classified as 'Inadequate'. Although this proportion has lessened considerably over the last decade, particularly in primary education, the association between schools' disadvantage contexts and students' educational outcomes remains worryingly strong and plays a key role when understanding 'failing' schools (Hutchinson 2016). If schools that are judged as Inadequate do not meet minimum standards after receiving a notice to improve, they can be closed down. In the cases of inadequacy, that action can be conversion or re-brokering a 'failing' school into an academy (Ofsted 2017).

What instruments are used to identify 'failing' schools in England? School Performance Category 'Inadequate' is based on inspectors' professional judgement of the school overall effectiveness. This judgement is informed by primary data collected by inspectors through inspection visits and second data analysis of academic performance (attainment and progress), and school and pupil contextual information summarised in the Inspection Data Summary Report (IDSR). The IDSR is a PDF file showing official national data for Ofsted inspectors to use when preparing for and during inspection. It aims to trigger inspectors' initial discussion with schools. The 2019 IDSR includes detailed information on pupils' academic performance (attainment and progress) as well as other critical measures (such as absences and exclusions). This statistical summary report also includes information regarding the context at the school level (such as percentage of Free School Meals, ethnicity, workforce, local area deprivation) Multi Academy Trust/Local Authority level (Ofsted grades profile), and year group level (attainment, context, progress trends). The aim of the IDSR is to support inspectors' professional judgement on the quality of a school by putting school data into context (i.e., highlighting quantitative trends and differences from national data), and supporting their interpretation through sentences called 'areas of interest' which describe qualitatively schools' historical data (DfE 2020). Therefore, inspectors judge the effectiveness of a school based on the analysis of secondary data, and data collected through inspection visits following the inspection framework. What remains implicit in this process is the weight that inspectors give to the different sources of evidence.

Cross-case comparison

The process of evaluating and constructing 'failing' schools as a result of accountability systems in Chile, the USA and England vary greatly. In line with Faubert's (2009) conclusions, we found that the methods, standards, data and instruments differ from system to system. However, the application of tools to measure the quality of education in the three countries identifies more and less effective schools, which in turn produces *winners* and *losers* in an accountability game. Thus, the current regulatory regimes make failure not only inevitable (Lefstein 2013), but also apparently desirable (Falabella 2020), as the three accountability systems align with the neoliberal market-oriented reforms previously described. This is further reflected in the fact that the three countries routinely implement *naming and shaming* strategies through public identification of 'failing' schools. Although many scholars have contested that these strategies lead to a spiral of decline following a

school being judged based on its performance on a narrow set of indicators, particularly for those schools positioned at the bottom of the pile (Stoll and Myers 2002; Perryman 2010), their voices tend to be ignored when assessing the impact of these policies.

Regarding who makes the decision about 'failing' schools and the system level each is located, key similarities emerged between Chile and England. Both rely on a powerful national institution that is autonomous from the central educational authority (ACE and Ofsted) to identify and classify 'failing' schools. The federal and state levels of the institutions that perform the same tasks in the USA, however, afford much more variation. Whilst the accountability systems converge at a national level in Chile and England, a more divergent and autonomous system is implemented in the USA.

Additionally, each system develops a different strategy when dealing with 'failing' schools. Whilst Chile and the USA contemplate a sequence of evaluation and support strategies oriented to enhance school improvement, England has progressively moved away from this model. Instead, 'Ofsted exists to be a force for improvement through intelligent, responsible and focused inspection and regulation' (Ofsted 2019, 4) that is expected to 'act as a trigger to others to take action' (Op. cit, 5). Unlike the cases of Chile and the USA, the English system is making clear that institutions other than Ofsted are responsible for implementing school improvement strategies.

Focusing on the instruments used to identify 'failing' schools, the three systems combine students' academic standardised tests with other non-academic indicators associated with school improvement. However, whilst England relies heavily on school inspection implemented by external inspectors before judging schools' effectiveness, targeted inspection visits are implemented only after the 'failing' school has been identified in Chile, relying exclusively on off-site analysis of secondary attainment and outcome data. In the USA, the extent to which inspections are conducted varies widely across states. Whilst the three systems consider schools' contextual factors, including location and student background characteristics, all mainly assign responsibility to schools for their low levels of student achievement (Bacchi 2000). Thus, little attention is given to the structures that create unequal starting points and subsequent outcomes. This seems to continue despite research stressing that differences in performance alone are incomplete reflections of school quality: factors such as regional, spatial, socioeconomic and other inequalities matter (Munoz-Chereau and Thomas 2016).

Discussion

Despite system differences, we conclude that the scope of the accountability mechanisms in place in the three cases is similar, as 'failing' schools are a by-product of neoliberal market-oriented reforms. The effects of the fabrication of *losers* in the accountability game is a very serious matter from social justice, conceptual and policy-oriented perspectives. Contrary to the expected policy theory, the labelling of schools fails to deliver on the promise of equal opportunities and non-discrimination of marginalised groups. For instance, the construction of 'failing' schools in the three accountability systems analysed above has seen increased segregation and inequality of opportunity between low-income and high-income students, hindering the distributive dimension of social justice (Flórez and Rozas 2020).

The emphasis on students' attainment and standardised educational outcomes ignores the social and cultural diversity of school communities, stigmatising schools serving more disadvantaged communities regardless of context (Gewirtz 2006). For accountability systems to seriously commit to distributional and recognitional dimensions of social justice would require a transition from standardisation to account for schools' context before attempting any meaningful comparisons between schools. By recognising that the relative influence of any school is by far smaller in magnitude than the relative importance of pupil intake and school context, the discussion would move towards the recognition of the challenges and strategies developed by schools working in challenging circumstances to deliver quality education (Tikly and Barrett 2011).

Focusing on the conceptual distinction, unless the judgement of school effectiveness is radically reconceptualised, schools positioned at the bottom of the pile will continue facing growing challenges and tough penalties, which in turn will exacerbate segregation and educational inequalities (Flórez and Rozas 2020). Moving away from deficit terminology by replacing it with the language of diversity and solidarity could impact positively the way these schools value and define themselves. However, although 'failing' schools are a minority in the three analysed accountability systems, they cannot be regarded as separate entities. Thus, it is not enough to conceptually change the label provided to the least effective schools, but to stop pretending that labelling them as 'failing' is a fair and valid practice.

Finally, whilst Chile, the USA and England's 'failing' schools represent a minority (less than 10%), they embody a powerful symbolic function for policy and all other schools. As Jones and Tymms (2014) explain, these schools represent 'the wrong side of the inspection thresholds' (322). If these schools share their common institutional struggles and ways in which learning and teaching respond to challenging contexts, a different positionality would emerge. By not being at the bottom, but at the top of their own game, the whole system could learn from their agency and empowerment, which in turn could impact positively their subjectivities and performance. These changes in policy would mean that instead of making individual schools responsible for their performance, responsibility would be shared at a local and national level which would allow to address the social inequalities that are strongly impacting the education delivered in schools.

References

*ACE. 2016a. *Construcción del Indicador Distribución por Niveles de Aprendizaje [Construction of the Distribution of Learning Levels Indicator]*. Santiago: Agencia de la Calidad de la Educación.

*ACE. 2016b. *Metodología de Construcción Indicadores Puntaje Simce y Progreso Simce [Methodology for the Construction of Simce Score and Simce Progress Indicators]*. Santiago: Agencia de la Calidad de la Educación.

*ACE. 2016c. *Metodología de Construcción Indicadores de Desarrollo de Personal y Social [Methodology for the Construction of Personal and Social Development Indicators]*. Santiago: Agencia de la Calidad de la Educación.

*ACE. 2017. *Panorama de La Gestión Escolar ¿Cómo Avanzamos En Calidad En Las Escuelas Que Más Apoyo Requieren? Primer Informe 2014-2015 [School management overview: How do we advance in quality in the schools that require the most support? First Report 2014-2015]*. Santiago, Chile: Agencia de la Calidad de la Educación.

*Ahumada Luis, Carmen Montecinos, and Álvaro González. 2012. "Quality Assurance in Chile's Municipal Schools: Facing the Challenge of Assuring and Improving Quality in Low Performing Schools." In *Quality Assurance and Management*, edited by Mehmet

Savsar, 183–192. Rijeka: IntechOpen.

*Aldeman, Chad, Anne Hyslop, Max Marchitello, Jennifer O'Neil Schiess, and Kaitlin Pennington. 2017. *An Independent Review of ESSA State Plans: Virginia.* Washington, D.C.: Bellwether Education Partners.

Altrichter, Herbert, and David Kemethofer. 2015. "Does Accountability Pressure through School Inspections Promote School Improvement?" *School Effectiveness and School Improvement* 26 (1): 32–56. doi:10.1080/09243453.2014.927369.

Bacchi, Carol. 2000. "Policy as Discourse: What Does It Mean? Where Does It Get Us?" *Discourse: Studies in the Cultural Politics of Education* 21 (1): 45–57. doi:10.1080/01596300050005493.

*Balfanz, Robert, Nettie Legters, Thomas C. West, and Lisa M. Weber. 2007. "Are NCLB's Measures, Incentives, and Improvement Strategies the Right Ones for the Nation's Low-Performing High Schools?" *American Educational Research Journal* 44 (3): 559–593. doi:10.3102/0002831207306768.

Ballou, Dale, and Matthew G. Springer. 2017. "Has NCLB Encouraged Educational Triage? Accountability and the Distribution of Achievement Gains." *Education Finance and Policy* 12 (1): 77–106. doi:10.1162/EDFP_a_00189.

Barber, Michael. 2002. "The Dark Side of the Moon: Imagining an End to Failure in Urban Education." In *No Quick Fixes: Perspectives on Schools in Difficulty*, edited by Louise Stoll and Kate Myers, 27–43. London: Routledge.

*Bellei, Cristián, and Xavier Vanni. 2015. "Chile: The Evolution of Educational Policy, 1980-2014." In *Education in South America: Education Around the World*, edited by Simon Schwartzman, 179–200. London: Bloomsbury Academic.

Brooks Hall, Jeffrey. 2017. "Examining school inspectors and education directors within the organisation of school inspection policy: perceptions and views" *Scandinavian Journal of Educational Research*, 61 (1): 112–126 doi:10.1080/00313831.2015.1120234.

Chapman, Christopher, and Alma Harris. 2004. "Improving Schools in Difficult and

Challenging Contexts: Strategies for Improvement." *Educational Research* 46 (3): 219–228. doi:10.1080/0013188042000277296.

*DfE. 2010. "Academies Act Notes". http://www.legislation.gov.uk/ukpga/2010/32/notes/contents

*DfE. 2016. "DfE Strategy 2015 to 2020: World-Class Education and Care". https://www.gov.uk/government/publications/dfe-strategy-2015-to-2020-world-class-education-and-care.

*DfE. 2020. "IDSR Guidance". https://www.gov.uk/guidance/school-inspection-data-summary-report-idsr-guide#overview-of-the-idsr

*Dragoset, Lisa, Jaime Thomas, Mariesa Hermann, John Deke, Susanne James-Burdumy et al. 2016. *Race to the Top: Implementation and Relationship to Student Outcomes: Executive Summary* (NCEE 2017-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

*Dragoset, Lisa, Jaime Thomas, Mariesa Hermann, John Deke, Susanne James-Burdumy et al. 2017. *School Improvement Grants: Implementation and Effectiveness: Executive Summary* (NCEE 2017-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Ehren, Melanie C. M., and Nichola Shackleton. 2016. "Mechanisms of Change in Dutch Inspected Schools: Comparing Schools in Different Inspection Treatments." *British Journal of Educational Studies* 64 (2): 185–213. doi:10.1080/00071005.2015.1019413.

Ehren, Melanie C. M., Johan E. Gustafsson, Herbert Altrichter, Guri Skedsmo, David Kemethofer, and Stefan G. Huber. 2015. "Comparing Effects and Side Effects of Different School Inspection Systems across Europe." *Comparative Education* 51 (3): 375–400. doi:10.1080/03050068.2015.1045769.

Falabella, Alejandra. 2014. "The Performing School: The Effects of Market & Accountability Policies." *Education Policy Analysis Archives* 22 (70): 1–29.

doi:10.14507/epaa.v22n70.2014.

*Falabella, Alejandra. 2020. “The Ethics of Competition: Accountability Policy Enactment in Chilean Schools’ Everyday Life.” *Journal of Education Policy* 35 (1): 23–45. doi:10.1080/02680939.2019.1635272.

Falabella, Alejandra, and Luis Felipe de la Vega. 2016. “Políticas de Responsabilización Por Desempeño Escolar: Un Debate a Partir de La Literatura Internacional y El Caso Chileno [Accountability Policies for School Performance: A Debate from International Literature and the Chilean Case].” *Estudios Pedagógicos* 42 (2): 395–413. doi:10.4067/S0718-07052016000200023.

Faubert, Violaine. 2009. *School Evaluation: Current Practices in OECD Countries and a Literature Review*. 42. OECD Education Working Paper. Paris. doi:10.1787/218816547156.

Flórez, María Teresa, and Tamara Rozas. 2020. “Accountability from a Social Justice Perspective: Criticism and Proposals.” *Journal of Educational Change* 21 (1): 157–182. doi:10.1007/s10833-019-09361-3.

Gewirtz, Sharon. 2006. Towards a Contextualized Analysis of Social Justice in Education. *Educational Philosophy and Theory* 38 (1): 69–81. https://doi.org/10.1111/j.1469-5812.2006.00175.x.

Gleeson, Denis, and Chris Husbands, eds. 2001. *The Performing School: Managing, Teaching, and Learning in a Performance Culture*. London: Routledge.

*Greany, Toby, and Rob Higham. 2018. *Hierarchy, Markets and Networks. Analysing the “Self-Improving School-Led System” Agenda in England and the Implications for Schools*. London: UCL Institute of Education Press.

Harvey, David. 2005. *A Brief History of Neoliberalism*. Oxford: Oxford University Press.

*Hurlburt, Steven, Kerstin Carlson Le Floch, Susan Bowles Therriault, and Susan Cole. 2011. *Baseline Analyses of SIG Applications and SIG-Eligible and SIG-Awarded Schools*. Washington, DC: National Center for Education Evaluation and Regional

Assistance, Institute of Education Sciences, U.S. Department of Education.

*Hutchinson, Jo. 2016. *School Inspection in England: Is There Room to Improve?* London: Education Policy Institute.

Jones, Karen, and Peter Tymms. 2014. "Ofsted's Role in Promoting School Improvement: The Mechanisms of the School Inspection System in England." *Oxford Review of Education* 40 (3): 315–330. doi:10.1080/03054985.2014.911726.

Lefstein, Adam. 2013. "The Regulation of Teaching as Symbolic Politics: Rituals of Order, Blame and Redemption." *Discourse: Studies in the Cultural Politics of Education* 34 (5): 643–659. doi:10.1080/01596306.2013.728361.

Levin, Henry., ed. 2018. *Privatizing Education: Can the School Marketplace Deliver Freedom of Choice, Efficiency, Equity, and Social Cohesion?.* New York: Routledge.

Luginbuhl, Rob, Dinand Webbink, and Inge de Wolf. 2009. "Do Inspections Improve Primary School Performance?" *Educational Evaluation and Policy Analysis* 31 (3): 221–237. doi:10.3102/0162373709338315.

Manna, Paul. 2010. *Collision Course: Federal Education Policy Meets State and Local Realities.* Washington, DC: CQ Press.

Maroy, Christian. 2009. "Convergences and Hybridization of Educational Policies around 'Post-bureaucratic' Models of Regulation." *Compare: A Journal of Comparative and International Education* 39 (1): 71–84. doi:10.1080/03057920801903472.

*Matthews, Felicity M. 2016. "Letting Go and Holding on: The Politics of Performance Management in the United Kingdom." *Public Policy and Administration* 31 (4): 303–323. doi:10.1177/0952076715615186.

Matthews, Peter, and Pam Sammons. 2005. "Survival of the Weakest: The Differential Improvement of Schools Causing Concern in England." *London Review of Education* 3 (2): 159–176. doi:10.1080/14748460500163989.

*McGuinn, Patrick. 2016. "From No Child Left behind to the Every Student Succeeds Act: Federalism and the Education Legacy of the Obama Administration." *Publius: The*

Journal of Federalism 46 (3): 392–415. doi:10.1093/publius/pjw014.

*Meckes, Lorena, and Rafael Carrasco. 2010. “Two Decades of SIMCE: An Overview of the National Assessment System in Chile.” *Assessment in Education: Principles, Policy & Practice* 17 (2): 233–248. doi:10.1080/09695941003696214.

Miles, Matthew B., A. M. Huberman, and Johnny Saldaña. 2014. *Qualitative Data Analysis. A Methods Sourcebook*. 3rd ed. Thousand Oaks, CA: SAGE Publications.

*MINEDUC. 2008a. “Separata: La Subvención Escolar Preferencial. Un Camino Para Mayor Equidad [Offprint: The Preferential School Subsidy. A Path for Increased Equity].” *Nuestros Temas* 29.

*MINEDUC. 2008b. “Separata: Manual Para La Elaboración Del Plan de Mejoramiento Educativo [Offprint: Manual for the Design of the Educational Improvement Plan].” *Nuestros Temas* 31.

*MINEDUC. 2015. *Plan de Evaluaciones Nacionales e Internacionales 2016-2020 [National and International Assessments Plan 2016-2020]*. Santiago, Chile: Unidad de Curriculum y Evaluación.

*Mintrop, Heinrich, and Tina Trujillo. 2005. “Corrective Action in Low Performing Schools: Lessons for NCLB Implementation from First-Generation Accountability Systems.” *Education Policy Analysis Archives* 13 (48): 1–27. doi:10.14507/epaa.v13n48.2005.

Munoz-Chereau, Bernardita and Sally Thomas. 2016. “Educational effectiveness in Chilean secondary education: comparing different ‘value added’ approaches to evaluate schools”. *Assessment in Education: Principles, Policy and Practice, 23 (1)*, 26-52. doi:10.1080/0969594X.2015.1066307.

Munter, Charles, Paul Cobb and Calli Shekell. 2016. “The Role of Program Theory in Evaluation Research: A Consideration of the What Works Clearinghouse Standards in the Cases of Mathematics Education.” *American Journal of Education* 37 (1): 7–26. doi:10.1177/1098214015571122.

OECD. 2013. *Synergies for Better Learning. An International Perspective on Evaluation and*

Assessment. Paris: OECD Publishing. doi:10.1787/9789264190658-en.

Office of Elementary & Secondary Education. 2020. *About Us.* Retrieved from https://oese.ed.gov/offices/office-of-administration/about-us/

*Ofsted. 2017. *Ofsted Strategy 2017–22*. London: Office for Standards in Education, Children's Services and Skills.

*Ofsted. 2019. *The Education Inspection Framework*. London: Office for Standards in Education, Children's Services and Skills.

Ozga, Jenny. 2016. "Trust in Numbers? Digital Education Governance and the Inspection Process." *European Educational Research Journal* 15 (1): 69–81. doi:10.1177/1474904115616629.

Penninckx, Maarten, Jan Vanhoof, Sven De Maeyer, and Peter Van Petegem. 2016. "Effects and Side Effects of Flemish School Inspection." *Educational Management Administration & Leadership* 44 (5): 728–744. doi:10.1177/1741143215570305.

*Penuel, William, Elizabeth Meyer, and Michelle Renée Valladares. 2016. *Making the Most of the Every Student Succeeds Act (ESSA): Helping States Focus on School Equity, Quality and Climate*. *National Education Policy Center*. Boulder, CO: National Education Policy Center.

Perryman, Jane. 2010. "Improvement after Inspection." *Improving Schools* 13 (2): 182–196. doi:10.1177/1365480210369878.

*Rentner, Diane S., Nancy Kober, and Matthew Frizzell. 2017. *Planning for Progress: States Reflect on Year One Implementation of ESSA*. Washington, DC: Center on Education Policy.

Rihoux, Benoît, and Bojana Lobe. 2009. "The Case for Qualitative Comparative Analysis (QCA): Adding Leverage for Thick Cross-Case Comparison." In *The SAGE Handbook of Case-Based Methods*, edited by David Byrne and Charles C Ragin, 222–242. London: SAGE Publications.

Rosenthal, Leslie. 2004. "Do School Inspections Improve School Quality? Ofsted Inspections

and School Examination Results in the UK." *Economics of Education Review* 23 (2). 143–151. doi:10.1016/S0272-7757(03)00081-5.

Rouse, Cecilia Elena, Jane Hannaway, Dan Goldhaber, and David Figlio. 2013. "Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure." *American Economic Journal: Economic Policy* 5, no. 2: 251-81.

Sahlberg, Pasi, and Jonathan Hasak. 2017. "Small Data for Big Change." *Education* 98 (1): 102.

Saw, Guan, Barbara Schneider, Kenneth Frank, I-Chien Chen, Venessa Keesler, and Joseph Martineau. 2017. "The impact of being labeled as a persistently lowest achieving school: Regression discontinuity evidence on consequential school labeling." American Journal of Education 123, no.: 585-613.

Schagen, Ian, and Sandie Schagen. 2003. "Analysis of National Value-Added Datasets to Assess the Impact of Selection on Pupil Performance." *British Educational Research Journal* 29 (4). 561–582. doi:10.1080/01411920301841.

Shaw, I., D. P. Newton, M. Aitkin, and R. Darnell. 2003. "Do OFSTED Inspections of Secondary Schools Make a Difference to GCSE Results?" *British Educational Research Journal* 29 (1). 63–75. doi:10.1080/0141192032000057375.

*Smith, Nelson, and Brandon Wright. 2017. *Leveraging ESSA to Support Quality-School Growth. Thomas B. Fordham Institute*. Washington, DC: Thomas B. Fordham Institute.

Stoll, Louise, and Kate Myers. 2002. "No Quick Fixes: Perspectives on Schools in Difficulties." London: Flamer.

Tikly, Leon, and Angela M. Barrett. 2011. "Social Justice, Capabilities and the Quality of Education in Low Income Countries." *International Journal of Educational Development* 31 (1). 3-14. doi: 10.1016/j.ijedudev.2010.06.001.

*United States Department of Education. 2020. *Office of Elementary & Secondary Education.* Washington, D.C.: Author. Retrieved from https://oese.ed.gov/

VanGronigen, Bryan A., and Coby V. Meyers. 2019. "How State Education Agencies Are

Administering School Turnaround Efforts: 15 Years after No Child Left Behind." *Educational Policy* 33 (3). 423-452. doi: 10.1177/0895904817691846.

*Virginia Department of Education. 2018. *Virginia Consolidated State ESSA Plan.* Richmond, VA: Author. Retrieved from https://www2.ed.gov/admins/lead/account/stateplan17/vaconsolidatedstateplanfinal.pdf