

Article manuscript

**Heidi Huilla**

Department of Education, University of Helsinki, Finland

P.O. Box 9, 00014 University of Helsinki, tel. +358-50-3385714,

heidi.huilla@helsinki.fi, ORCID iD: 0000-0002-2845-3496

# **A Circle of Research on Disadvantaged Schools, Improvement and Test-Based Accountability**

## **Abstract**

This study analyses how studies on disadvantaged schools, improvement and test-based accountability relate to each other. The analysis covers 69 studies on disadvantaged schools reported in prestigious educational journals and conducted in 1995–2015. Educational policies related to evaluation and accountability define the official goals of schooling, and the aim in this article is to analyse how the chosen studies discuss these educational policies and understand school success and failure. The following questions were asked: What typologies related to test-based accountability can be constructed in research on disadvantaged schools? What understandings of good schools are embedded in the identified typologies? Disadvantaged schools are at the centre of improvement and therefore also the target of evaluative policy practices. The results show that research supports test-based accountability practices, and that critical studies on school improvement are in the minority.

## **Keywords:**

Disadvantaged schools, school improvement, failing schools, accountability

## **Introduction**

This study analyses the relations among studies on disadvantaged schools, school improvement and test-based accountability practices. The focus is on how a good school is constructed in the interrelations between studies and hard accountability practices. The

data consists of 69 English-language articles reporting research on disadvantaged schools and published in 1995–2015 in six prestigious educational journals. The data was identified following a systematic literature review and subjected to qualitative content analysis.

Educational policies define the aim and targets of schooling. Many education systems have adopted accountability practices utilising hard techniques such as the standardised testing of student outcomes to control the achievement of these policy aims. Various studies have raised critical discussions on how test-based accountability, along with large-scale assessments such as the Programme for International Student Assessment (PISA), narrow educational goals. Evaluation based on standardised high-stakes testing outcomes in particular has caused concern. Schools have a lot at stake in achieving acceptable outcome levels, and this has had profound consequences on everyday life and understanding related to schooling (Ball, 2003; Fullan, 2011; Hamre, Morin & Ydesen, 2018; Kauko, Rinne & Takala, 2018; Kelly, 2018; Lindblad, Pettersson & Popkewitz, 2018; Lingard et al., 2016; Mons 2009; Ozga, Dahler-Larsen, Segerholm & Simola, 2011; Ranson, 2003; Ravitch, 2010; Sahlberg, 2011; Thrupp, 1999; Wrigley, 2003, 2011; Whitty, 2002).

Previous research has shown that schools with a somewhat marginalised student body have difficulty in achieving success measured in terms of learning outcomes (Coleman et al., 1966; Muijs et al., 2004). Consequently, so-called ‘disadvantaged schools’ tend to be taken as targets of improvement practices to produce better outcomes. As a concept, disadvantage(d schools) is context related and it cannot be universally defined. It commonly refers to a school with a student body that is predominantly low in terms of socio-economic status. Disadvantaged schools in the current study are understood through the lenses of the studies that were chosen for the analysis, which could be classified as studies on disadvantaged schools based on database indexes.

The research objective of this study is to explore how academic studies on disadvantaged schools are constructing and possibly maintaining an understanding of the perceived necessity for test-based accountability and related improvement practices in the process of achieving the politically determined aims of schooling. A further aim is to analyse how a good school is fabricated in the process, in other words the role of educational policies in the definition of successful and failing schools in educational research. This matter has previously raised intense discussions among scholars (Cuban, 2003; Gewirtz, 1998; Reay, 2004; Slee & Weiner, 2001; Teddlie & Reynolds, 2001; Thrupp, 2001; Thrupp, Lauder & Robinson, 2002; Townsend, 2001; Wrigley, 2003, 2011). The research questions addressed are: What typologies related to test-based accountability can be constructed in research on disadvantaged schools? What understandings of good schools are embedded in the identified typologies?

The argumentation in this paper proceeds as follows. Previous research on evaluation politics, and especially on test-based accountability, is reviewed next, and after that the history, political connections and formation of quality assurance and evaluation policies are discussed, as are various critical comments on the problems involved and the alternatives to test-based accountability. Then the current study is presented in terms of aims, data, methods and analysis. The results are presented through the two main

typologies identified. The concluding discussion assesses the critical implications of the findings for future research.

## An Intensified Testing Culture as a Means of School Improvement: Background

The national testing of pupil achievement became more widespread in Europe during the 1990s. Since then, and especially in England, tests have been established more fundamentally to monitor and improve the quality of education, and to make educational systems more effective. (Eurydice, 2009.) This follows the trend in the United States and Australia, for example, as well as New Zealand, which nevertheless changed course and discontinued using standardised assessments rather recently (see Thrupp, 2018). The political, societal and theoretical background behind the spreading of testing regimes can be traced to and understood from different but related perspectives. One consequence of globalisation is the accelerating convergence in educational policies worldwide. Included in this convergence of educational systems is the ‘governance turn’: the concept of governance or steering from a distance is used to conceptualise the increased potential for monitoring and self-control arising from the explosion in the production of numerical data. (Lindblad et al., 2018; Ozga et al., 2011; Simola, Ozga, Segerholm, Varjo & Normann Andersen, 2011; see also Kauko, Takala & Rinne, 2018.) Tests influence actions on the grass-roots level and then provide information to the top level, thereby creating a circle of quantified data and social control. International large-scale assessments (LSAs) such as the Programme for International Student Assessment (PISA) tend to affect practices in national contexts, thereby enhancing test-based accountability (Thrupp, 2018).

The neoliberal ethos of competition and the global turn towards new public management have served to mediate the new knowledge-governance relationship, and perhaps also vice versa (Gunter, Grimaldi, Hall & Serpieri, 2016; Simola et al., 2011; Thrupp, 2018). According to Natalie Mons (2009), the theoretical background of standardised assessment comprises New Public Management (NPM) and political evaluation on the macro-level, and the economics of education as well as research on school effectiveness on the micro-level. NPM, political evaluation and the economics of education highlight the cost-benefit ratio of education, pupil achievement being understood as the quantified output of resources invested in public-sector education. The research on school effectiveness connects rather easily with this theoretical background, given the focus on the organisational and strategic structures within schools and the outcomes they produce (see also Thrupp, 1999; Wrigley, 2003).

Standardised assessment practices appear to be rational and objective, but their connection to these specific schools of thought makes them political (Ball, 2018; Lindblad et al., 2018). National assessment practices should be perceived as political and as parts of larger shifts in educational politics related to processes of globalisation and convergence, and to advanced technologies in quality assurance and evaluation (QAE) (Ozga et al., 2011; see also Ozga, Seddon & Popkewitz, 2006). These larger shifts have been referred to as

*'policy or education by the numbers'* (Lingard et al., 2016, p.1), *'the infrastructure of accountability'* (p.3) and *'a comparativistic paradigm'* (Lindblad et al., 2018, p. 5).

Pasi Sahlberg (2011a, 2011b), building on the work of Andy Hargreaves, gives a theoretical and critical summary of recent developments in educational improvement in terms of the Global Educational Reform Movement (GERM) (or GERMS, if one considers the path-dependent vernacular versions - see Lingard et al., 2016, p. 6), which is spreading like a virus, whereas Michael Fullan (2001) writes about wrong drivers in educational reforms. Decision makers in many countries have resorted to hard QAE techniques, meaning national testing, ranking lists, inspection for evaluation, benchmarking and national goal setting, in order to improve education (Gray et al., 2011). According to GERM, six globally common features are perceptible in attempts to improve school quality, and student achievement in particular: standardisation, an increased focus on core subjects such as literacy and numeracy as well as predetermined results in the curriculum, the transfer of models from the corporate world, increased control, and high-stakes accountability policies (Sahlberg, 2011a, 2011b).

## Problems with and alternatives to test-based accountability

Improvement through hard QAE, and more specifically test-based accountability, provoke severe criticism and arouse concerns among scholars. First, there seems to be no empirical consensus with regard to testing, not even for the alleviation of educational inequalities (Mons, 2009). Second, the relationship between testing and what is considered to be education's core purpose appears to be contested: testing conflicts with inclusive education, social justice, democracy, equal opportunities and recognition, and may even be counterproductive to these aims (Bingham, 2001; Hamre et al., 2018). Thus the question arises about the goals of education. *'Should education be a production factor in the achievement of economic prosperity or should education serve personal development values, Bildung, and making students capable of leading rich and fulfilling lives?'* (Hamre et al., 2018, p. 254; see also Ball, 2018; Kauko, Takala et al., 2018; Lingard et al., 2016).

The accountability system, which is based on high-stakes testing, seems to be a double-edged sword for disadvantaged schools. On the one hand, given that schools with a disadvantaged student body are most likely to 'fail' (e.g. Muijs et al., 2004), the idea behind demanding better results from these schools is to prevent the staff from using the socio-economic background of pupils as an excuse, and to make sure that disadvantaged pupils have equal chances to undertake further studies. On the other hand, the assumption that the success of a school equates with good outcomes in standardised high-stakes tests requires the staff of disadvantaged schools to excel personally on a daily basis, and overshadows other aims that schooling might have. (Stahl, 2017; Whitty, 2002.)

Apart from narrowing educational aims, hard accountability measures have had other consequences. As the concept implies, pupils, teachers and schools have a lot at stake. For pupils such measures determine their school career. Permanently excluding pupils from schools also seems to be connected with hard accountability (Daniels, 2018). As for

teachers, pupil success and failure are tied to rewards and punishments related to salary, for example, whereas schools failing tests mean special measures for improvement or closing down. High-stakes tests seem to evoke the ‘teaching to the test’ phenomenon, and the offering of support to those who are close to achieving an acceptable result. Tests emphasise individualism and encourage individuals to aim at ‘*self-optimization*’ (Hamre et al., 2018, p. 257), which also leaves the responsibility for success or failure to the individual. All in all, the chosen models shape the identities and thinking of teachers and students, which in turn shape practices in schools and the overall ethos of education (Ball, 2003; Ranson, 2003).

Alternatives to GERM have also been explored extensively in the research literature. What is common to them all is that they require a larger shift in the converging neoliberal ethos of (educational) politics. More effort could be put into redistributing wealth and funding more justly. Schools could focus on values such as democratic education and teach more complex skills such as problem-solving and social skills instead of a couple of core subjects. (Lupton, 2005; Mons, 2009; Olssen, Codd & O’Neill, 2004; Paille, 2013; Thrupp, 1999; Whitty, 2002; Wrigley, 2003.) The staff should be allowed to take risks: the emphasis should be on high professionalism and intrinsic motivation, and therefore responsibility would occur as a consequence (Fielding, 2001; Sahlberg, 2011a; Fullan, 2011). Hard QAE practices seem to be at odds with these proposals (see also Gray et al., 2011). Alternatively, accountability could be context-related, used for developmental purposes, and it should benefit grass-roots-level work instead of acting as a control device (Kauko, Rinne et al., 2018; Lingard et al., 2016).

## The study

The debate, especially on test-based accountability, is not only between critical researchers and policy practitioners, but also between research disciplines. Critics claim that studies emphasising standardised school efficiency have managed to create a discourse that dominates current educational practices (Cuban, 2003; Gewirtz, 1998; Reay, 2004; Thrupp, Lauder & Robinson, 2002; Wrigley, 2003, 2011). The testing culture has indeed changed the nature of educational research, and even of what counts as educational research (Lingard et al., 2016). Recent studies in the European context have identified the governing mechanisms through which actors on different levels influence educational research: evaluation systems, competition and economic competitiveness have an effect not only on the grass-roots level, but also on research (Powell, Zapp, Marques & Biesta, 2018; Zapp et al., 2018).

This study is an analysis of literature reporting research on disadvantaged schools in 1995–2015. The focus of the analysis is on how academic research literature discusses the relationship between test-based accountability policies and disadvantaged schools, and on the role of hard accountability practices in the definition of what constitutes a good school: policies define the goals that schooling has, but what is the role of research in this? The concept of a good school is understood as politically articulated and entangled in

historical and discursive formations, which makes “success” a social phenomenon. As a consequence, studies on disadvantaged schools are understood as being involved in creating and maintaining the ethos behind school improvement (see Thrupp, 1999; Wrigley, 2003, 2011). Disadvantaged schools are often targeted for improvement. It is therefore necessary to take a critical view on how the desired quality is understood: if it is understood in terms of outcomes in standardised tests it might not, in reality, improve the lives of marginalised students (Whitty, 2002; Ravitch, 2010).

To capture the relationship between research and test-based accountability practices the following questions are addressed. What typologies related to test-based accountability can be constructed in the research on disadvantaged schools? What understandings of good schools are embedded in the identified typologies?

The data consist of 69 English-language articles published in six prestigious educational journals, identified following a systematic literature review (Petticrew & Robert, 2006) and analysed by means of qualitative content analysis (Schreier, 2012). The articles: 1) relate to research on disadvantaged schools; 2) focus on primary or lower-secondary schools; 3) are peer-reviewed; and 4) were published between 1995 and 2015. These twenty years were chosen because it is a controllable time span and still manages to describe the near history of research on disadvantaged schools. It also covers the rise of the test-based accountability period. The focus was on primary and lower-secondary levels because practically all children go through these educational stages. The data collection proceeded as follows. First, the search was targeted on databases containing educational studies. Within ProQuest and EBSCOHOST the search was targeted at Academic Search Complete, Education Research Complete, SocINDEX with Full Text, ERIC, PsycINFO and Sociological Abstracts databases, as well as SCOPUS. The following Boolean algorithm was applied:

(disadvantaged OR deprived OR "low socioeconomic" OR "low socio economic" OR poor OR failing OR marginali\* OR low-performing OR underachieving OR “under achieving”) NEAR/3 schools

OR

challenging NEAR/1 schools.

The search was further targeted on abstracts or ‘all but full texts’<sup>1</sup>. The first extraction limited the number of articles. All the abstracts were read through, and the articles that did not meet the selection criteria were discarded. After this, two criteria drove the choice of journals: they contained several of the studies that were chosen based on the abstracts, and had a high impact factor (IF, Journal Citation Reports, year 2015). Journals that engaged in discussions about how disadvantaged schools should be studied were also emphasised in the selection (see e.g. Downey & Gondron, 2016; Slee & Weiner, 2001;

---

<sup>1</sup> If choosing ‘abstract’ as the target of the search was not possible, ‘all but full texts’ targeted the search to abstracts. I did not want to target the search to the full texts, because disadvantage might obviously be mentioned somewhere in the paper even if it is not in the focus of the study. I wanted to find articles that actually studied ‘disadvantaged’ schools.

Teddle & Reynolds, 2001; Thrupp, 2001; Townsend, 2001). The six journals chosen were the American Educational Research Journal (IF 2.924), the British Educational Research Journal (IF 1.124), Educational Evaluation and Policy Analysis (IF 2.02), the Journal of Educational Policy (IF 2.174), School Effectiveness and School Improvement (IF 1.333), and Sociology of Education (IF 2,000). They contained all in all 69 studies on disadvantaged schools. It is worth taking a critical look at the journals and articles this enquiry reaches. The ranking of journals and their being considered prestigious based on a numerical value assigned to them could be perceived as a circle maintaining itself: publishing in restricted journals is desirable, and this consequently strengthens the position of these prestigious journals. However, the fact that the journals chosen for this study are likely to have a major impact on further research justifies the analysis of articles contained in them.

The theory-related qualitative content analysis (Schreier, 2012) proceeded as follows. It was targeted at the introduction and conclusion of the articles in question. If there was a discussion, it was also included. First, all the articles were read through. Next, the introduction, conclusion and possible discussion were extracted and fed into ATLAS.ti, scientific software for qualitative analysis. All references to neoliberalism, (social)justice, (in)equality, community, outcomes (all related expressions), purpose/aim, success, quality, fail(ure), improvement, effective(ness), and learning were coded, and the excerpts were extracted from each article. The excerpts were re-read, and the most relevant ones were gathered and collated into a table. The perspectives in each article were then summarised critically in line with the theoretical focus of this study. This enabled the construction of typologies from each article, and of two main typologies. The main typologies could be understood as analytical constructions: although not all the articles are clearly one or the other, they all have traits of either one of the extremes. The process was hardly straightforward, which is typical for qualitative analysis. The reading and understanding proceeded as a dialogue, not only with the theoretical part but also with the data itself: there were 69 studies on disadvantaged schools and it was natural to use them to so as to understand what matters when studying disadvantaged schools. The sharpening of the research and the analytical focus was a backwards-and-forwards process. The data was perused several times, and a final round was executed to increase the reliability of the analysis.

## Results

In answer to the first research question of what typologies related to test-based accountability can be constructed in the research on disadvantaged schools, two main typologies were identified in the data: the aim in the first one is to improve disadvantaged schools to achieve better high-stakes outcomes, and the second one severely criticises current policies.

According to the effectiveness or quality-as-numbers typology, success and failure are based on national educational policies. Limits of goodness and badness are determined by the respective national accountability system, in other words the standardised high-stakes testing carried out on the school level<sup>2</sup>. Twenty-four articles fall clearly to this typology, and a further 11 consider contextual factors, but do not question high-stakes testing outcomes as the goal of improvement<sup>3</sup>. The research settings of these studies reflect hard accountability practices, and the aim of many is to inform policymakers: school success is straightforwardly attributed as success in measurable outcomes.

At the other extreme, 16 of the 69 studies openly refer to schools as institutions for developing more than academic learning outcomes and cognitive skills<sup>4</sup>. Most of them are critical of test-based accountability structures and their tendency to narrow down the purpose of schooling<sup>5</sup>. Instead, they demand quality that is beyond measurable test outcomes<sup>6</sup>, a focus on motivation, engagement<sup>7</sup> and intersectional troubles in learning<sup>8</sup>, practices and structures that promote social justice and social inclusion<sup>9</sup>, the social formation of pupils as citizens<sup>10</sup> and the development of a democratic character<sup>11</sup>. School quality relates to an internalised understanding of democratic societies. These 16 studies also depict schools as part of a bigger entity trying to affect disadvantage. Consequently ‘succeeding schools’ are part of larger themes of social justice.

The following paragraphs describe the two main typologies and how they relate to accountability policy and practices, improvement goals and perceived goodness, in other words quality that becomes constructed through different understandings of the goals. The results are presented through examples given the extensive data and space restrictions. This description also addresses the second research question concerning the various understandings of good schools embedded in the identified typologies. A ‘good school’ means different things depending on the improvement goal: it may mean good individuals in schools managing to achieve good test outcomes, for example, or it could imply a

---

<sup>2</sup> Elliot, Arthurs and Williams 2000; Langer 2000; 2001; Riley, Burrell, and McCallum 2004; Nye, Konstantopoulos, and Hedges 2004; van de Grift and Houtveen 2006; Lauen 2007; Ylimaki, Jacobson, and Drysdale 2007; Lleras 2008; Gross, Booker, and Goldhaber 2009; Donaldson and Moore Johanson 2010; Loeb, Kalogrides, and Lai Horng 2010; Ronfeldt 2012; Corcoran Schwartz and Weinstein 2012; Lee Lauen and Gaddis 2012; Zimmer and Guarino 2013; Carlson, Cowen and Fleming 2013; Cohen-Vogel, Feng and Osborne-Lampkin 2013; Hart 2014; Anzia and Moe 2014; Gershenson and Langbein 2015; Heers et al. 2015; McEwan et al. 2015; Xu, Özek, and Hansen 2015.

<sup>3</sup> Mussoline and Shouse 2001; Woods and Levačić 2002; Nicolaidou and Ainscow 2005; Dworkin 2005; Harris et al. 2006; Borman and Dowling 2006; van den Berg and van Reekum 2011; Ingersoll and May 2012; McNaughton, Lai, and Hsiao 2012; Ronfeldt, Loeb, and Wyckoff 2013; Jennings and Sohn 2014.

<sup>4</sup> Richie 2004; Milbourne 2004; Gewirtz et al. 2005; Mintrop and Trujillo 2007; Luyten, Peschar, and Coe 2008; Araújo 2009; Power and Frandji 2010; Lupton and Hempel-Jorgensen 2012; Crossouard 2012; Angus 2012; Riley 2013; Hernandez-Martinez and Williams 2013; Thomas 2013; Smyth and McInerney 2014; Singh, Heimans, and Glasswell 2014; Trujillo and Woulfin 2014.

<sup>5</sup> Mintrop and Trujillo 2007; Luyten, Peschar, and Coe 2008; Araújo 2009; Power and Franji 2010; Lupton and Hempel-Jorgensen 2012; Crossouard 2012; Angus 2012; Hernandez-Martinez and Williams 2013; Thomas 2013; Trujillo and Woulfin 2014.

<sup>6</sup> Mintrop and Trujillo 2007; Power and Frandji 2010; Hernandez-Martinez and Williams 2013.

<sup>7</sup> Luyten, Peschar, and Coe 2008.

<sup>8</sup> Crossouard 2012.

<sup>9</sup> Lupton and Hempel-Jorgensen 2012; Angus 2012; Thomas 2013.

<sup>10</sup> Araújo 2009.

<sup>11</sup> Trujillo and Woulfin 2014.

complex interplay among structural-level solutions, families, students and schools. This is illustrated through examples.

Most of the studies that use standardised test outcomes as empirical data or in descriptions of the features of case schools build their research setting so as to improve the results. In other words, school success is straightforwardly attributed in these studies<sup>12</sup>: schools and the people within them succeed by producing good-enough learning outcomes, and what is under scrutiny is whether they succeed in this, and if so, how. Such an approach assigns specific meanings to some expressions: ‘making a difference’ equals producing maximised test-based learning outcomes; ‘effective’ equals high-performing, while the low-performing are less effective; and ‘performing’ refers to test-based outcomes, for example. As Anzia and Moe (2014, p.84) write, *‘these are the schools in greatest need of improvement, and thus in greatest need of high quality teachers’*: improvement and quality here refer to *performance*, which refers to test outcomes. Carlson et al. (2013, p.179) ‘[–] show that low-performing voucher students tend to move from the voucher sector into lower performing and less effective public schools than the typical public school student attends, whereas high-performing students transfer to better public schools’. This quotation illustrates how “better” refers to schools performing well in test outcomes.

The performance of individuals becomes a major importance and is in focus when the goal is to produce good test outcomes. This creates ‘kinds of people’ (see Popkewitz, 2018) in schools: teachers, principals and students. Good teachers are created through an internalised appreciation of the importance of children’s succeeding in tests. Nye, Konstantopoulos and Hedges (2004) and Xu, Özek and Hansen (2015), for example, suggest tackling inequality by allocating ‘*effective*’ teachers to disadvantaged schools. Teacher turnover is understood as a means of improving ineffective schools. Ylimaki, Jacobson and Drysdale (2007, p.378) describe the aspired principal of disadvantaged schools as ‘[having traits of] *persistence, empathy, passion, and flexible, creative thinking*’ -- [and having] *empathy for the barriers to learning that poverty can produce, [but not allowing] these conditions to be used as excuses for poor performance*’. A picture of a good pupil or rather a bad one is also constructed: disadvantaged pupils are at-risk because they do not perform as well as their more advantaged peers, but disadvantage associates only with poor outcomes from the school’s perspective. If schools were to emphasise good behaviour, emotional control and better learning outcomes they might make these children into better learners in terms of test-based accountability. Inequality is understood as unfairness in school outcomes.

There are also studies that fall in between the two main typologies. Teacher turnover as a means of school improvement is considered problematic given its consequences beyond numbers to remaining teachers, for example (Ronfeldt, Loeb & Wyckoff, 2013; see also Ingersoll & May, 2012). These “in-between” studies aim at good

---

<sup>12</sup> Elliot, Arthurs and Williams 2000; Langer 2000; 2001; Riley, Burrell, and McCallum 2004; Nye, Konstantopoulos, and Hedges 2004; Borman and Dowling 2006; van de Grift and Houtveen 2006; Lauen 2007; Ylimaki, Jacobson, and Drysdale 2007; Lleras 2008; Gross, Booker, and Goldhaber 2009; Donaldson and Moore Johanson 2010; Loeb, Kalogrides, and Lai Horng 2010; Ronfeldt 2012; Corcoran Schwartz and Weinstein 2012; Lee Lauen and Gaddis 2012; Zimmer and Guarino 2013; Carlson, Cowen and Fleming 2013; Cohen-Vogel, Feng and Osborne-Lampkin 2013; Hart 2014; Anzia and Moe 2014; Gershenson and Langbein 2015; Heers et al. 2015; McEwan et al. 2015; Xu, Özek, and Hansen 2015.

outcomes but not regardless of context,<sup>13</sup> and therefore have similar traits as those promoting the view that the wider social and political context is inseparable from school functioning, and that the focus should be on system-level, long-term and deeper-rooted problems of social inequality instead of only on individual actors in schools<sup>14</sup>. Merry (2013; see also Chudgar & Luschei, 2009) investigated inequality using Large Scale Assessments (LSAs) and test outcomes as data. According to the results of the tests and the LSAs, the differences in achievement between the US and Canada were attributable to social conditions rather than the ineffectiveness of the US school system. The author calls for contextualisation in research, and also shows how LSAs and tests can be used to make contextualised mechanisms of inequality visible (see also Butler et al., 2007). Downey et al. (2008) measured school effectiveness by analysing the effects on learning based on times when children are not at school. They claim that a good school cannot be judged on test achievements, and that schools should be evaluated based on their added value (see also Borman & Dowling, 2006). Value-added league tables are also criticised in the data, however: *‘[–] the need to distinguish ‘external’ and ‘internal’ variables [is] a distinction which is ontologically confused and empirically elusive’* (Power and Frandji, 2010, p.391).

Many of the articles representing the typology that takes a critical stand on testing argue against the notion of school improvement based solely on test outcomes. Hard quality-assurance and evaluation practices in particular are fiercely criticised<sup>15</sup>.

‘A single, externally imposed definition of which knowledge is worth knowing, where expertise lies, and what a good school looks like becomes foregrounded in these settings; learning rooted in students’ background and culture becomes secondary, at best.’

(Trujillo & Woulfin, 2014, p.288.)

These researchers openly consider, and call for politicians to consider, the aim and purpose of schooling, and juxtapose (measurable) cognitive, academic achievements and other aims in education<sup>16</sup>. Understanding schools as feeders for economic competition is also subjected to criticism (Lupton and Hempel-Jorgensen, 2012; Trujillo & Woulfin, 2014). As Angus (2012) states, for example, neoliberal policies and policy as numbers conflict with social justice, economic opportunity and democratic outcomes. Araújo (2009) calls for *‘becoming a person’* for democratic societies as the outcome of schooling instead of measurable test-based outcomes. The aim of Gewirtz, Dickson, Power, Halpin and Whitty (2005), Lupton and Hempel-Jorgensen (2012), and Power and Frandji (2010) is to shed

---

<sup>13</sup> Mussoline and Shouse 2001; Woods and Levačić 2002; Nicolaidou and Ainscow 2005; Dworkin 2005; Harris et al. 2006; Borman and Dowling 2006; van den Berg and van Reekum 2011; Ingersoll and May 2012; McNaughton, Lai, and Hsiao 2012; Ronfeldt, Loeb, and Wyckoff 2013; Jennings and Sohn 2014.

<sup>14</sup> Warren 1996; Milbourne 2004; Gorard 2005; Downey, von Hippel, and Hughes 2008; Gordon 2008; Araújo 2009; Chudgar and Luschei 2009; Power and Frandji 2010; Thrupp and Lupton 2011; Gorard 2012; Lupton and Thrupp 2013; Merry 2013.

<sup>15</sup> Milbourne 2004; Ritchie 2004; Mintrop and Trujillo 2007; Luyten, Peschar, and Coe 2008; Araújo 2009; Power and Franji 2010; Lupton and Hempel-Jorgensen 2012; Angus 2012; Crossouard 2012; Thomas 2013; Lupton and Thrupp 2013; Hernandez-Martinez and Williams 2013; Riley 2013; Singh, Heimans, and Glasswell 2014; Trujillo and Woulfin 2014.

<sup>16</sup> Mintrop and Trujillo 2007; Luyten, Peschar, and Coe 2008; Araújo 2009; Angus 2012; Lupton and Hempel-Jorgensen 2012; Riley 2013; Thomas 2013; Trujillo and Woulfin 2014.

light on structural, system-level problems through analyses of recognition and recognitive justice (see also Thomas 2013). They suggest that politics and research should focus both on pedagogies that make a difference and on practices that tackle inequality on the system level.

Some of these studies shed critical light on policies aimed at helping disadvantaged pupils and show that the policies face (unexpected) problems: the practices seem to be controversial in the sense that they may inflict damage on and exclude disadvantaged children, even though the aim is the opposite<sup>17</sup>. As Smith and Stovall (2008) state: *'strategies to redevelop poorly performing 'inner city' schools and the poor neighborhoods where they are usually located has real potential to do more harm than good for the families the policy is supposed to benefit'*. According to Stevens and van Houtte (2011), evaluation systems have an effect on the teacher-pupil relationship, and hard accountability systems do not function for the benefit of the worst performers. Other ways of improving disadvantaged schools and lives are suggested, such as fostering an understanding of local contextual conditions to support families and communities,<sup>18</sup> as well as reshaping structures to create more equal societies. Success becomes constructed as the complex interplay between relationships within schools and policies that work on the level of families.

The concepts of quality and success are also under critical scrutiny in the "critical typology". Hard QAE narrows the possibilities for innovative work in schools (Milbourne, 2004). It is claimed, for example, that it restricts not only other taught subjects but also subjects that are measured in terms of tests, such as mathematics (Hernandez-Martinez & Williams, 2013). This resonates with Luyten, Peschar and Coe (2008), who express concern that engaged and motivated reading might, in fact, conflict with performance requirements. Hard QAE narrows the understanding of quality: Mintrop and Trujillo (2007) write about their research schools, which were surprisingly similar even though some of them were succeeding in terms of measured outcomes and others were not. The writers conclude that the 'succeeding' schools took the accountability system more seriously and aimed to succeed in that manner, whereas the others also emphasised other initiatives and aspects of schooling (see also Riley, 2013). These examples reveal how policies affect the way quality and success become understood and internalised within schools, and how these internalisations differ: it is not necessarily the case that one school is good and another one bad, it is more a question of colliding understandings of quality (Mintrop & Trujillo, 2014). Singh, Heimans and Glasswell (2014, p. 838) describe these colliding understandings as *'collateral realities'* and claim that researchers cannot separate themselves from them: when they are doing their work they are creating reality, an understanding of what good teaching, effective schooling and quality learning are.

---

<sup>17</sup> Richie 2004; Gewirtz et al. 2005; Mintrop and Trujillo 2007; Luyten, Peschar, and Coe 2008; Smith and Stovall 2008; Stevens and van Houtte 2011; Smyth and McInerney 2014; Trujillo and Woulfin 2014.

<sup>18</sup> Gewirtz et al. 2005; Gordon 2008; Angus 2012; Riley 2013; Trujillo & Woulfin 2014.

## Conclusion

This study addressed two questions: what typologies related to test-based accountability can be constructed in research on disadvantaged schools, and what understandings of good schools are embedded in them. It became evident that the outcomes of standardised, high-stakes tests play a major role in defining what schools and people within them should be aiming for. They also construct a picture of a good school and a good human being within a school through success or failure in test outcomes. Current policies affect the themes and execution of research, often avoiding open or embedded criticism. To consider this connection critically: what kind of research is left undone or is not accepted for publication in journals that are considered the most prestigious? Educational politics and research resemble a circle in which good and bad schooling are defined in terms of how well pupils do in standardised high-stakes tests, after which the focus turns to how bad schools and the people in them could succeed better, and this is then monitored on the same scale. It is impossible to improve schools in other directions within this circle. Practices that emphasise hard quality assurance and evaluation remain the dominant practice of educational policy, fuelled and maintained by educational research. Kauko, Rinne et al. (2018, 183) describe the phenomenon as follows: [The results of this project --] *'indicate that standardised testing feeds a need for more testing (Piattoeva & Saari 2018) and that quality becomes simultaneously a means of problematizing education and providing a solution for it (Minina et al. 2018).'*

Does this really improve disadvantaged lives? Concentrating on outcomes highlights a rather mechanistic view of people, their relationships and the work they are doing. When people are evaluated only through the lens of outcomes, little room is left for human error, caring beyond effective results and stability. An American researcher and former advocate of standardised high-stakes testing, Diane Ravitch, claims (2010) that schools cannot improve unless poverty is tackled and unless the goal of schooling, what education is for, is (re)defined. This follows the conclusion in many of the articles analysed for this study. Thomas S. Popkewitz (2018, p. 231) claims that *'the paradox of the international comparisons is its inscription of difference that 'makes' differences so that some can never be at the "top"'*. If testing and comparisons indeed maintain the division between 'good and bad', and the bad are usually the disadvantaged, and if performativity pressures therefore do create exclusion and do more harm than good, as some of the data articles argue<sup>19</sup>, this is the phenomenon on which research could shed more light.

Disadvantaged people might benefit from a more critical perspective in the research on school improvement, including studies on mechanisms of socioeconomic inequality (e.g. Chudgar & Luschei, 2009; Merry, 2013), recognition (e.g. Gewirtz et al., 2005; see also Bingham, 2001; Fraser, 2009) and socially just pedagogies (e.g. Lupton & Hempel-Jorgensen, 2012). Both pedagogies that make a difference and practices that tackle inequality on the system level could be examined (Lupton & Hempel-Jorgensen, 2012), but such studies are clearly in the minority. Sixteen of the 69 studies comprising the data openly demand wider discussion on the purposes of schooling, and radical change

---

<sup>19</sup> see Gewirtz et al. 2005; Luyten, et al. 2008; Mintrop and Trujillo, 2007; Richie, 2004; Smith and Stovall, 2008; Smyth and McInerney, 2014; Stevens and van Houtte, 2011; Trujillo and Woulfin, 2014

especially in how school success is understood nowadays. They also consider the wider social and political context of disadvantaged schools. If the aimed-at success were understood as forging a connection between a human, society and moral, it would also place more comprehensive demands on accountability systems (Autio, 2014; Fullan, 2011; Sahlberg, 2011), but it might put quality back on the agenda instead of emphasising the measuring in itself (see Kauko, Rinne et al., 2018). As Lingard et al. (2016) conclude: ‘*We also argue that we cannot reject the need for accountability in education; rather, what we need to do is reconceptualize it, so that systems and schools are held accountable for their educative and social justice purposes, but in ways that are productive, democratic, and socially just.*’ (p. 15) ‘*Importantly, accountability systems must operate differently in different contexts*’ (p. 154). The results of this study emphasise the need to put complexity back on the agenda. In all probability there is not one simple measurable solution to problems concerning disadvantaged schools and the people within and around them.

## Acknowledgements

I am grateful to the anonymous referees and the journal editor for their valuable comments, which were perceptive and improved this article considerably. I also thank Assistant Professor Sonja Kosunen, Doctor Helena Rajakaltio and Professor Piia Seppänen for their comments. Doctor Hannele Pitkänen gave me invaluable help in the final phases.

## Funding

This work was supported by the Jenny and Antti Wihuri foundation; and the University of Helsinki.

## References

- Autio, T. (2014) The Internationalization of Curriculum Studies. Introductory Chapter One. In W. Pinar, (Ed.), *International Handbook of Curriculum Research. 2.nd edition*. New York, Routledge.
- Ball, S. J. (2003) The Teacher's Soul and the Terrors of Performativity. *Journal of Education Policy*, 18 (2), 215–228. DOI: 10.1080/0268093022000043065

- Ball, S. J. (2018) The banality of numbers. In Hamre, B., Morin, A. & Ydesen, C. (Eds.). *Testing and inclusive schooling: International challenges and opportunities* (pp. 79–86). Milton Park, Abingdon, Oxon; New York, NY: Routledge.
- Bingham, C. (2001). *Schools of Recognition, Identity Politics and Classroom Practices*. Lanham: Rowman & Littlefield publishers, INC.
- Bourdieu, P. (1986). The Forms of Capital. In Ball, S. J. (2004). *The RoutledgeFalmer reader in sociology of education* (pp. 15–29). London: Taylor & Francis.
- Coleman, J. S., Cambell, E. A., Hobson, C. J., McPortland, J., Mood, A. M., Weinfeld, F. D. & York, R. L. (1966). *Equality of Educational Opportunity*. Washington DC: US Government Printing Office.
- Crozier, G. (2015). Middle-Class Privilege and Education. *British Journal of Sociology of Education* 36 (7), 1115–1123. DOI: 10.1080/01425692.2015.1076249
- Cuban, L. (2003). *Why is it so hard to get good schools?* New York: Teachers College Press.
- Downey, D. B. & Condron, D. J. (2016). Fifty Years since the Coleman Report: Rethinking the Relationship between Schools and Inequality. *Sociology of Education*, 89 (3), 207–220. DOI: 10.1177/0038040716651676
- Eurydice, (2009). National Testing of Pupils in Europe: Objectives, Organisation and Use of Results. Brussels: Eurydice.
- Fielding, M. (2001). OFSTED, Inspection and the Betrayal of Democracy. *Journal of Philosophy of Education*, 35 (4), 695–709. DOI: 10.1111/1467-9752.00254
- Fraser, N. (2009). *Scales of Justice: Reimagining Political Space in a Globalizing World*. New York: Columbia University Press.
- Fullan, M. (2011). Choosing the Wrong Drivers for Whole System Reform. *Seminar Series 204, Melbourne: Centre for Strategic Education*. Retrieved from: <https://michaelfullan.ca/wp-content/uploads/2016/06/13396088160.pdf>
- Gewirtz, S. (1998). Can All Schools be Successful? An exploration of the Determinants of School ‘Success’. *Oxford Review of Education* 24 (4), 439–457. DOI: 10.1080/0305498980240402
- Gray, J., Croxford, L., Strømbæk Pedersen, C., Rinne, R., Silmäri-Salo, S., Simola, H., and Mäkinen-Streng, M. (2011). Teachers’ perceptions of quality assurance and evaluation. In Ozga, J., Dahler-Larsen, P., Segerholm, C. & Simola, H. (Eds.) *Fabricating Quality in Education, Data and Governance in Europe*. New York, Routledge.
- Gunter, H., Grimaldi, E., Hall, D. & Serpieri, R. (2016). *New public management and the reform of education*. Abingdon, Oxon; New York, NY: Routledge.
- Hamre, B., Morin, A. & Ydesen, C. (2018). Optimizing the educational subject between testing and inclusion in an era of neoliberalism, musings on a research agenda and its future perspectives. In Hamre, B., Morin, A. & Ydesen, C. (Eds.). *Testing and inclusive schooling: International challenges and opportunities* (pp. 254–261). Milton Park, Abingdon, Oxon; New York, NY: Routledge.
- Journal Citation Reports, (2015). Retrieved from: <https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action>
- Kauko, J., Rinne, R. & Takala, T. (2018). Conclusion. In Kauko, J., Rinne, R. & Takala, T. (Eds.) *Politics of Quality in Education, A Comparative Study of Brazil, China,*

- and Russia* (pp.180–190). Milton Park, Abingdon, Oxon; New York, NY: Routledge.
- Kauko, J., Takala, T. & Rinne, R. (2018). Comparing politics of quality in education. In Kauko, J., Rinne, R. & Takala, T. (Eds.) *Politics of Quality in Education, A Comparative Study of Brazil, China, and Russia* (pp.1–17). Milton Park, Abingdon, Oxon; New York, NY: Routledge.
- Kelly, P. (2018). Standardized assessment and the shaping of neoliberal student subjectivities. In Hamre, B., Morin, A. & Ydesen, C. (Eds.). *Testing and inclusive schooling: International challenges and opportunities* (pp.64–78). Milton Park, Abingdon, Oxon; New York, NY: Routledge.
- Lindblad, S., Pettersson, D. & Popkewitz, T. S. (2018). Getting the numbers right, an introduction. In Lindblad, S., Pettersson, D. & Popkewitz, T. S. (Eds.). *Education by the numbers and the making of society: The expertise of international assessments* (pp. 1–20). New York, NY: Routledge.
- Lingard, B., Martino, W., Rezai-Rashti, G. & Sellar, S. (2016). *Globalizing educational accountabilities*. New York; London: Routledge.
- Lupton, R. (2005). Social Justice and School Improvement: Improving the Quality of Schooling in the Poorest Neighbourhoods. *British Educational Research Journal* 31 (5), 589–604. DOI: 10.1080/0141192050024075
- Muijs, D., Harris, A., Chapman, C., Stoll, L. & Russ, J. (2004) Improving Schools in Socioeconomically Disadvantaged Areas - A Review of Research Evidence. *School Effectiveness and School Improvement*, 15 (2), 149–175. <http://dx.doi.org/10.1076/sesi.15.2.149.30433>
- Mons, N. (2009) Theoretical and Real Effects of Standardised Assessment. Eurydice network. Retrieved from: [http://eacea.ec.europa.eu/education/eurydice/documents/thematic\\_reports/111EN.pdf](http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/111EN.pdf)
- Olssen, M., Codd, J. & O’Neill, A-M. (2004). *Education Policy: Globalization, Citizenship and Democracy*. London: Sage publications.
- Ozga, J., Dahler-Larsen, P., Segerholm, C. & Simola, H. (2011). *Fabricating Quality in Education, Data and Governance in Europe*. New York: Routledge.
- Ozga, J., Seddon, T., & Popkewitz, T. S. (2006). *Education research and policy: steering the knowledge-based economy*. New York: Routledge.
- Paulle, B. (2013). *Toxic Schools, High-Poverty Education in New York and Amsterdam*. Chicago: The University of Chicago Press.
- Powell, J. J. W., Zapp, M., Marques, M. & Biesta, G. (2018). Introduction, (Re)Constructing institutional change in European educational research. In Zapp, M., Marques, M. & Powell, J. J. W. with contributions by Biesta, G. & Helgetun, J. B. (Eds.), *European educational research (re)constructed: Institutional change in Germany, the United Kingdom, Norway, and the European Union* (pp. 7–22). Didcot, Oxford: Symposium Books.
- Petticrew, M. & Robert, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford: Blackwell.
- Popkewitz, T. S. (2018). Anticipating the future society, The cultural inscription of numbers and international large-scale assessment. In Lindblad, S., Pettersson, D. &

- Popkewitz, T. S. (Eds.), *Education by the Numbers and the Making of Society, the Expertise of International Assessments* (pp. 222–238). New York, NY: Routledge.
- Ranson S. (2003). Public accountability and the age of neo-liberal governance. In Lingard, B. & Ozga, J. (Eds.) *The RoutledgeFalmer reader in education policy and politics* (pp. 198–219). London: Routledge.
- Ravitch, D. (2010). *The Death and Life of the Great American School System, How Testing and Choice Are Undermining Education*. New York: Basic Books.
- Reay, D. (2004). “Mostly Roughs and Toughs”: Social Class, Race and Representation in Inner City Schooling. *Sociology* 38 (5), 1005–1023. DOI: 10.1177/0038038504047183
- Sahlberg, P. (2011a). *Finnish Lessons, What can the world learn from educational change in Finland?* New York: Teachers College Press.
- Sahlberg, P. (2011b). The Fourth Way of Finland. *Journal of Educational Change*, 12(2), 173–185. DOI 10.1007/s10833-011-9157-y
- Schreier, M. (2012). *Qualitative Content Analysis in Practice*. London: SAGE Publications.
- Simola, H., Ozga, J., Segerholm, C., Varjo, J. & Normann Andersen, V. (2011). Governing by numbers. In Ozga, J., Dahler-Larsen, P., Segerholm, C. & Simola, H. (Eds.). *Fabricating Quality in Education, Data and Governance in Europe* (96–106). New York: Routledge.
- Slee, R. & Weiner, G. (2001). Education Reform and Reconstruction as a Challenge to Research Genres: Reconsidering School Effectiveness Research and Inclusive Schooling. *School Effectiveness and School Improvement*, 12 (1), 83–98. <http://dx.doi.org/10.1076/sesi.12.1.83.3463>
- Stahl, G. (2017). *Ethnography of a neoliberal school: Building cultures of success* (1st edition.). New York: Routledge.
- Teddle, C. & Reynolds, D. (2001). Countering the Critics: Responses to Recent Criticisms of School Effectiveness Research, *School Effectiveness and School Improvement*, 12 (1), 41–82. <http://dx.doi.org/10.1076/sesi.12.1.41.3458>
- Thrupp, M. (1999). *Schools Making a Difference. Let's Be Realistic!* Buckingham: Open University Press.
- Thrupp, M. (2018). *The search for better educational standards: A cautionary tale*. Cham: Springer.
- Thrupp, M., Lauder, H. & Robinson, T. (2002). School Composition and Peer Effects. *International Journal of Educational Research*, 37 (5), 483–504. [https://doi.org/10.1016/S0883-0355\(03\)00016-8](https://doi.org/10.1016/S0883-0355(03)00016-8)
- Townsend, T. (2001). Satan or Saviour? An Analysis of Two Decades of School Effectiveness Research. *School Effectiveness and School Improvement*, 12 (1), 115–129. <http://dx.doi.org/10.1076/sesi.12.1.115.3461>
- Whitty, G. (2002). *Making Sense of Education Policy: Studies in the Sociology and Politics of Education*. London: Paul Chapman Publishing.
- Wrigley, T. (2003). *Schools of Hope. A New Agenda for School Improvement*. Trentham: Stoke-on-Trent.
- Wrigley, T. (2011). Paradigms of School Change. *Management in education*, 25 (2), 62–66. DOI: 10.1177/0892020611398929

Zapp, M., Marques, M. & Powell, J. J. W. (2018). On the future of educational research (governance). In Zapp, M., Marques, M. & Powell, J. J. W. with contributions by Biesta, G. & Helgetun, J. B. (Eds.). *European educational research (re)constructed: Institutional change in Germany, the United Kingdom, Norway, and the European Union* (pp. 189–200). Didcot, Oxford: Symposium Books.

## Data articles:

- Alexander, K. L., Entwisle, D. R. & Olson, L. S. (2001). Schools, Achievement, and Inequality: A Seasonal Perspective. *Educational Evaluation and Policy Analysis*, 23 (2), 171–191. <https://doi.org/10.3102/01623737023002171>
- Angus, L. (2012). Teaching within and against the Circle of Privilege: Reforming Teachers, Reforming Schools. *Journal of Education Policy*, 27 (2), 231–251. DOI:10.1080/02680939.2011.598240
- Anzia, S. F. & Moe, T. M. (2014). Collective Bargaining, Transfer Rights, and Disadvantaged Schools. *Educational Evaluation and Policy Analysis*, 36 (1), 83–111. <https://doi.org/10.3102/0162373713500524>
- Araújo, M. (2009). A Fresh Start for a ‘Failing School’? A Qualitative Study. *British Educational Research Journal*, 35 (4), 599–617. DOI: 10.1080/01411920802642439
- Borman, G. D. & Dowling, N. M. (2006). Longitudinal Achievement Effects of Multiyear Summer School: Evidence from the Teach Baltimore Randomized Field Trial. *Educational Evaluation and Policy Analysis*, 28 (1), 25–48. <https://doi.org/10.3102/01623737028001025>
- Butler, T., Hamnett, C., Ramsden, M. & Webber, R. (2007). The Best, the Worst and the Average: Secondary School Choice and Education Performance in East London. *Journal of Education Policy*, 22 (1), 7–29. DOI: 10.1080/02680930601065718
- Carlson, D., Cowen, J. M. & Fleming, D. J. (2013). Life After Vouchers: What Happens to Students Who Leave Private Schools for the Traditional Public Sector? *Educational Evaluation and Policy Analysis*, 35 (2), 179–199. <https://doi.org/10.3102/0162373712461852>
- Charlton, E., Mills, A., Martino, W. & Beckett, L. (2005). Sacrificial Girls: A Case Study of the Impact of Streaming and Setting on Gender Reform. *British Educational Research Journal*, 33 (4), 459–478. DOI: 10.1080/01411920701434011
- Chingos, M. M. & West, M. R. (2015). The Uneven Performance of Arizona’s Charter Schools. *Educational Evaluation and Policy Analysis*, 37 (1), 120–134. <https://doi.org/10.3102/0162373715576077>
- Chudgar, A. & Luschei, T. L. (2009). National Income, Income Inequality, and the Importance of Schools: A Hierarchical Cross-National Comparison. *American Educational Research Journal*, 46 (3), 626–658. <https://doi.org/10.3102/0002831209340043>
- Cohen-Vogel, L. (2011). ”Staffing to the Test”: Are Today’s School Personnel Practices Evidence Based? *Educational Evaluation and Policy Analysis*, 33 (4), 483–505. <https://doi.org/10.3102/0162373711419845>

- Cohen-Vogel, L., Feng, L. & Osborne-Lampkin, L. (2013). Seniority Provisions in Collective Bargaining Agreements and the “Teacher Quality Gap”. *Educational Evaluation and Policy Analysis*, 35 (3), 324–343. <https://doi.org/10.3102/0162373713482765>
- Corcoran, S. P., Schwartz, A. E. & Weinstein, M. (2012). Training Your Own: The Impact of New York City’s Aspiring Principals Program on Student Achievement. *Educational Evaluation and Policy Analysis*, 34 (2), 232–253. <https://doi.org/10.3102/0162373712437206>
- Crossouard, B. (2012). Absent Presences: The Recognition of Social Class and Gender Dimensions within Peer Assessment Interactions. *British Educational Research Journal*, 38 (5), 731–748. DOI: 10.1080/01411926.2011.580047
- Elliot, J., Arthurs, J. & Williams, R. (2000). Volunteer Support in the Primary Classroom: The Long-Term Impact of One Initiative upon Children’s Reading Performance. *British Educational Research Journal*, 26 (2), 227–244. <http://dx.doi.org/10.1080/01411920050000962>
- Donaldson, M. L. & Moore Johanson, S. (2010). The Price of Missassignment: The Role of Teaching Assignments in Teach for America Teachers’ Exit from Low-Income Schools and the Teaching Profession. *Educational Evaluation and Policy Analysis*, 32 (2), 299–323. <https://doi.org/10.3102/0162373710367680>
- Downey, D. B., von Hippel, P. T. & Hughes, M. (2008). Are “Failing” Schools Really Failing? Using Seasonal Comparison to Evaluate School Effectiveness. *Sociology of Education*, 81 (3), 242–270. <https://doi.org/10.1177/003804070808100302>
- Dworkin, G. (2005). The No Child Left Behind Act: Accountability, High-Stakes Testing, and Roles for Sociologists. *Sociology of Education*, 78 (2), 170–174. <http://dx.doi.org/10.1177/003804070507800205>
- Gershenson, S. & Langbein, L. (2015). The Effect of Primary School Size on Academic Achievement. *Educational Evaluation and Policy Analysis*, 37 (1S), 135S–155S. <https://doi.org/10.3102/0162373715576075>
- Gewirtz, S., Dickson, M., Power, S., Halpin, D. & Whitty, G. (2005). The Deployment of Social Capital Theory in Educational Policy and Provision: The Case of Education Action Zones in England. *British Educational Research Journal*, 31 (6), 651–673. DOI: 10.1080/01411920500314620
- Gorard, S. (2005). Academies as the “Future of Schooling”: Is This an Evidence-Based Policy? *Journal of Education Policy*, 20 (3), 369–377. <http://dx.doi.org/10.1080/02680930500117321>
- Gorard, S. (2012) Who Is Eligible for Free School Meals? Characterising Free School Meals as a Measure of Disadvantage in England. *British Educational Research Journal*, 38 (6), 1003–1017. DOI: 10.1080/01411926.2011.608118
- Gorard, S., Taylor, C. & Fitz, J. (2002). Does School Choice Lead to “Spirals of Decline”? *Journal of Education Policy*, 17 (2), 367–384. DOI: 10.1080/02680930210127612
- Gordon, J. (2008). Community Responsive Schools, Mixed Housing, and Community Regeneration. *Journal of Education Policy*, 23 (2), 181–192. DOI: 10.1080/02680930701853237
- Gross, B., Booker, T. K. & Goldhaber, D. (2009). Boosting Student Achievement: The Effect of Comprehensive School Reform on Student Achievement. *Educational*

- Evaluation and Policy Analysis*, 31 (2), 111–126. <https://doi.org/10.3102/0162373709333886>
- Harris, A., Chapman, C., Mujis, D., Russ, J. & Stoll, L. (2006). Improving Schools in Challenging Contexts: Exploring the Possible. *School Effectiveness and School Improvement*, 17 (4), 409–424. DOI: 10.1080/09243450600743483
- Harris, D. & Williams, J. (2012). The Association of Classroom Interactions, Year Group and Social Class. *British Educational Research Journal*, 38 (3), 373–397. DOI: 10.1080/01411926.2010.548547
- Hart, C. M. D. (2014). Contexts Matter: Selection in Means-Tested School Voucher Programs. *Educational Evaluation and Policy Analysis*, 36 (2), 186–206. <https://doi.org/10.3102/0162373713506039>
- Heers, M., Ghysels, J., Groot, W. & van der Brink, H. M. (2015). Differentiated Effects of Community Schooling on Cognitive and Socio-Emotional Learning Outcomes. *School Effectiveness and School Improvement*, 26 (3), 354–381. DOI: 10.1080/09243453.2014.975138
- Hernandez-Martinez, P. & Williams, J. (2013). Against the Odds: Resilience in Mathematics Students in Transition. *British Educational Research Journal*, 39 (1), 44–59. DOI: 10.1080/01411926.2011.623153
- Ingersoll, R. M. & May, H. (2012). The Magnitude, Destinations, and Determinants of Mathematics and Science Teacher Turnover. *Educational Evaluation and Policy Analysis*, 34 (4), 435–464. <https://doi.org/10.3102/0162373712454326>
- Jennings, J. & Sohn, H. (2014). Measure for Measure: How Proficiency-Based Accountability Systems Affect Inequality in Academic Achievement. *Sociology of Education*, 87 (2), 125–141. doi: 10.1177/0038040714525787
- Langer, J. A. (2000). Excellence in English in Middle and High School: How Teachers' Professional Lives Support Student Achievement. *American Educational Research Journal*, 37 (2), 397–439. <https://doi.org/10.3102/00028312037002397>
- Langer, J.A. (2001). Beating the Odds: Teaching Middle and High School Students to Read and Write Well. *American Educational Research Journal*, 38 (4), 837–880. <https://doi.org/10.3102/00028312038004837>
- Lauen, D. L. (2007). Contextual Explanations of School Choice. *Sociology of Education*, 80 (3), 179–209. <https://doi.org/10.1177/003804070708000301>
- Lauen, D. L. & Gaddis, S. M. (2012). Shining a Light or Fumbling in the Dark? The Effects of NCLB's Subgroup-Specific Accountability on Student Achievement. *Educational Evaluation and Policy Analysis*, 34 (2), 185–208. <https://doi.org/10.3102/0162373711429989>
- Lleras, C. (2008). Race, Racial Concentration, and the Dynamics of Educational Inequality across Urban and Suburban Schools. *American Educational Research Journal*, 45 (4), 886–912. <https://doi.org/10.3102/0002831208316323>
- Loeb, S., Kalogrides, D. & Lai Horng, E. (2010). Principal Preferences and the Uneven Distribution of Principals across Schools. *Educational Evaluation and Policy Analysis*, 32 (2), 205–229. <https://doi.org/10.3102/0162373710369833>
- Lupton, R. & Hempel-Jorgensen, A-J. (2012). The Importance of Teaching: Pedagogical Constraints and Possibilities in Working-Class Schools. *Journal of Education Policy*, 27 (5), 601–620. DOI: 10.1080/02680939.2012.710016

- Lupton, R. & Thrupp, M. (2013). Headteachers' Readings of and Responses to Disadvantaged Contexts: Evidence from English Primary Schools. *British Educational Research Journal*, 39 (4), 769–788. DOI: 10.1080/01411926.2012.683771
- Luyten H., Peschar J. & Coe R. (2008). Effects of Schooling on Reading Performance, Reading Engagement, and Reading Activities of 15-Year-Olds in England. *American Educational Research Journal*, 45 (2), 319–342. <https://doi.org/10.3102/0002831207313345>
- McEwan, P. J., Murphy-Graham, E., Torres Iribarra, D. Aguilar, C. & Rápalo, R. (2015). Improving Middle School Quality in Poor Countries: Evidence from the Honduran Sistema de Aprendizaje Tutorial. *Educational Evaluation and Policy Analysis*, 37 (1), 113–137. <https://doi.org/10.3102/0162373714527786>
- McNaughton, S., Lai, M. K. & Hsiao, S. (2012). Testing the Effectiveness of an Intervention Model Based on Data Use: A Replication Series across Clusters of Schools. *School Effectiveness and School Improvement*, 23 (2), 203–228. DOI: 10.1080/09243453.2011.652126
- Merry, J. J. (2013). Tracing the U.S. Deficit in PISA Reading Skills to Early Childhood: Evidence from the United States and Canada. *Sociology of Education*, 86 (3), 234–252. <https://doi.org/10.1177/0038040712472913>
- Milbourne, L. (2004). Children, Families and Inter-Agency Work: Experiences of Partnership Work in Primary Education Settings. *British Educational Research Journal*, 31 (6), 675–695. DOI: 10.1080/01411920500314653
- Mintrop, H. & Trujillo, T. (2007). The Practical Relevance of Accountability Systems for School Improvement: A Descriptive Analysis of California Schools. *Educational Evaluation and Policy Analysis*, 29 (4), 319–352. <https://doi.org/10.3102/0162373707309219>
- Mussoline, L. J. & Shouse, R. C. (2001). School Restructuring as a Policy Agenda: Why One Size May Not Fit All. *Sociology of Education*, 74 (1), 44–58. URL: <http://www.jstor.org/stable/2673144>
- Nicolaidou, M. & Ainscow, M. (2005). Understanding Failing Schools: Perspectives from the Inside. *School Effectiveness and School Improvement*, 16 (3), 229–248. DOI: 10.1080/09243450500113647
- Nordstrum, L. E. (2012). Incentives to Exclude: The Political Economy Constraining School Fee Abolition in South Africa. *Journal of Education Policy*, 27 (1), 67–88. DOI: 10.1080/02680939.2011.604138
- Nye, B., Konstantopoulos, S. & Hedges, L. V. (2004). How Large Are Teacher Effects? *Educational Evaluation and Policy Analysis*, 26 (3), 237–257. <https://doi.org/10.3102/01623737026003237>
- Power, S. & Frandji, D. (2010). Education Markets, the New Politics of Recognition and the Increasing Fatalism towards Inequality. *Journal of Education Policy*, 25 (3), 385–396. <http://dx.doi.org/10.1080/02680930903576404>
- Riley, J., Burrell, A. & McCallum, B. (2004). Developing the Spoken Language Skills of Reception Class Children in Two Multicultural, Inner-City Primary Schools. *British Educational Research Journal*, 30 (5), 657–672. DOI: 10.1080/0141192042000234638

- Riley, K. A. (2013). Walking the Leadership Tightrope: Building Community Cohesiveness and Social Capital in Schools in Highly Disadvantaged Urban Communities. *British Educational Research Journal*, 39 (2), 266–286. DOI: 10.1080/01411926.2012.658021
- Ritchie, G. (2004). Quantifying the Effects of Teacher Movements Between Schools in New Zealand: To Schools that Hath, Shall Be Given. *Journal of Education Policy*, 19 (1), 57–79. DOI: 10.1080/0268093042000182636
- Ronfeldt, M. (2012). Where Should Teachers Learn to Teach? Effects of Field Placement School Characteristics on Teacher Retention and Effectiveness. *Educational Evaluation and Policy Analysis*, 34 (1), 2–26. <https://doi.org/10.3102/0162373711420865>
- Ronfeldt, M., Loeb S. & Wyckoff J. (2013). How Teacher Turnover Harms Student Achievement. *American Educational Research Journal*, 50 (1), 4–36. <https://doi.org/10.3102/0002831212463813>
- Singh, P., Heimans, S. & Glasswell, K. (2014). Policy Enactment, Context and Performativity: Ontological Politics and Researching Australian National Partnership Policies. *Journal of Education Policy*, 29 (6), 826–844. DOI: 10.1080/02680939.2014.891763
- Smith, J. & Stovall, D. (2008). 'Coming Home' to New Homes and New Schools: Critical Race Theory and the New Politics of Containment. *Journal of Education Policy*, 23 (2), 135–152. DOI: 10.1080/02680930701853062
- Smyth, J. & McInerney, P. (2014). 'Ordinary Kids' Navigating Geographies of Educational Opportunity in the Context of an Australian 'Place-Based Intervention'. *Journal of Education Policy*, 29 (3), 285–301. DOI: 10.1080/02680939.2013.794304
- Stevens, P. A. J. & van Houtte, M. (2011). Adapting to the System or the Student? Exploring Teacher Adaptations to Disadvantaged Students in an English and a Belgian Secondary School. *Educational Evaluation and Policy Analysis*, 33 (1), 59–75. <https://doi.org/10.3102/0162373710377112>
- Thomas, G. (2013). A Review of Thinking and Research about Inclusive Education Policy, with Suggestions for a New Kind of Inclusive Thinking. *British Educational Research Journal*, 39 (3), 473–490. DOI: 10.1080/01411926.2011.652070
- Thrupp, M. & Lupton, R. (2011). Variations on a Middle Class Theme: English Primary Schools in Socially Advantaged Contexts. *Journal of Education Policy*, 26 (2), 289–312. DOI: 10.1080/02680939.2010.508178
- Trujillo, T.M. & Woulfin S.L. (2014). Equity-Oriented Reform Amid Standards-Based Accountability: A Qualitative Comparative Analysis of an Intermediary's Instructional Practices. *American Educational Research Journal*, 51 (2), 253–293. <https://doi.org/10.3102/0002831214527335>
- van de Grift, W. J. C. M., & Houtveen, A. A. M. (2006). Underperformance in Primary Schools. *School Effectiveness and School Improvement*, 17 (3), 255–273. DOI: 10.1080/09243450600697317
- van der Berg, M. & van Reekum, R. (2011). Parent Involvement as Professionalization: Professionals' Struggle for Power in Dutch Urban Deprived Areas. *Journal of Education Policy*, 26 (3), 415–430. DOI: 10.1080/02680939.2010.543155

- Warren, J. R. (1996). Educational Inequality among White and Mexican-Origin Adolescents in the American Southwest: 1990. *Sociology of Education*, 69 (2), 142–158. DOI: 10.2307/2112803
- Woods, P. A. & Levačić, R. (2002). Raising School Performance in the League Tables (Part 2): Barriers to Responsiveness in Three Disadvantaged Schools. *British Educational Research Journal*, 28 (2), 227–247. DOI: 10.1080/01411920120122167
- Ylimaki, R. M., Jacobson, S. L. & Drysdale, L. (2007). Making Difference in Challenging, High-Poverty Schools: Successful Principals in the USA, England, and Australia. *School Effectiveness and School Improvement*, 18 (4), 361–381. DOI: 10.1080/09243450701712486
- Xu, Z., Özek, U. & Hansen, M. (2015). Teacher Performance Trajectories in High- and Low-Poverty Schools. *Educational Evaluation and Policy Analysis*, 37 (4), 458–477. DOI: 10.3102/0162373714561513
- Zimmer, R. W. & Guarino, C. M. (2013). Is there Empirical Evidence that Charter Schools "Push Out" Low-Performing Students? *Educational Evaluation and Policy Analysis*, 35 (4), 461-480. <https://doi.org/10.3102/0162373713498465>