

School Evaluation or Disciplinary Subjection? A consideration of the ‘complementary’ evaluation of New Zealand’s Education Review Office

LEON BENADE

School of Education, Auckland University of Technology

Abstract

Educational evaluation in New Zealand schools harmonises with the reformist agenda of public choice theory. Variously accused of engaging in politics of blame or teacher bashing, ERO has attempted to present itself as supporting schools in a cycle of ‘complementary’ review, which synthesises external accountability with internal learning from review. This paper will examine the underpinning epistemology of the Education Review Office approach to review and reflect on its particular research method, arguing that it is motivated by a commitment to ‘evidence-led’ teaching, a problematic concept. Questions are raised regarding the ‘complementary’ nature of the review process to establish whether there is commitment to democratic participation by schools in the review process, as implied in the concept of complementarity. With reference to Foucauldian concepts, it will be argued that the so-called ‘complementary’ dimension of the Education Review Office process is characterised by the administration of technologies of self-discipline and self-punishment, and that ‘learning’ in this context is deeply punitive in nature.

Keywords: evaluation; evidence-led teaching; complementary review; accountability

Introduction

The Education Review Office (ERO) was established by the Education Act of 1989, one development amongst many in the post-1984 establishment of the neoliberal state in New Zealand. This reform period in New Zealand witnessed the marketisation of schools, emphasising individual choice over the community (Codd, 2005; Gordon, 1997; McKenzie, 1997; Olssen & Morris Matthews, 1997; Snook, 2003). An important rationale for the existence of ERO is neoliberal consumer choice, which demands the transparent provision of information about schools to enable parents to make well-informed choices. It pays to understand too that the reformist climate in which ERO was born has its source in global developments, which includes the role of global governance (such as UNESCO, World Bank and Organisation for Economic Cooperation and Development) in proposing particular directions for education reform (Dale, 1999; Dale & Robertson, 2002). One such suggestion is the role of data in driving practice, latterly toward the notion of ‘value-added’ (personal excellence for the individual student, rather than competitive ranking across students) (Martinic, 2012).

ERO is a government department empowered to visit schools and early childhood centres to conduct reviews (for the purposes of this article, the focus will be on state schools only). It reports back to the board of trustees and community of the school as well as to the Minister with responsibility for ERO (Education Review Office, n.d. c). ERO has sweeping ‘powers of entry and inspection’ under Section 28, Part 327 of the Education Act (1989) (New Zealand Government, 2009). ERO reports are public record and available to the public, either in hard copy from ERO itself (or the relevant school) or on the ERO website (www.ero.govt.nz). These factors have contributed to the sense that ERO ‘hunts down’ failing or struggling schools (Thrupp, 1997), participates in the ‘politics of blame’ (1998) and engages in ‘teacher bashing’ (Benade, 2009). Thrupp has argued that ERO has latched onto the school improvement and teacher effectiveness literature (2008) to justify its particular accountability regime over schools and teachers.

In recent times, ERO has shifted its position by engaging in more streamlined and flexible evaluation reviews. Simultaneously, schools are now invited to play a greater role in their own internal evaluation, to complement the external evaluation conducted by ERO. In this article, I will set out to highlight the relationship between evaluation and research, showing that despite similarities, evaluation serves a quite different purpose. However, given the similarity, it is instructive to consider whether ERO demonstrates an evident theoretical approach in its documentation, and what its dominant methodological practice consists in (setting aside for now the problem of the link between methodology and method). I will suggest, by way of critique, the concept of evaluation underpinning the work of ERO continues to be based on low trust accountability consistent with notions related to neoliberalism. Specifically, ERO seeks to coordinate its evaluation process with an epistemology based on the discourse of school effectiveness. The notion of complementary review, despite its attempt to engage schools in democratic dialogue and participation, in reality disguises structures that act as technologies by which boards exercise both self-discipline and self-punishment in advance of, during, and after, ERO review visits.

Evaluation

‘The Education Review Office (ERO) is the New Zealand government department that evaluates and reports on the education and care of students in schools and early childhood services’ (Education Review Office, n.d. a, ‘About Us’). The work that ERO therefore conducts is, in essence, evaluation, which by definition, entails some form of measurement or judgement against a set of prescribed descriptors or performance criteria (more of which shortly). The field of evaluation is one that bears resemblance to many aspects of the work conducted by researchers in social fields or disciplines, such as education or sociology (Scriven, 1996), yet it seems those engaged professionally in evaluation have on-going debates in regard to the relationship between theory and method (Kushner, 2002; 2005; Scriven, 1996), such as finding the balance (Scriven, 1996) or confusing methods (tools of analysis) with methodology (‘the logic of enquiry rather than its technology’ (Kushner, 2002, p. 252). Scriven, in contributing to this debate almost two decades ago (by which time ERO was well established) tellingly argued across these debates: ‘The time has come to realize that we now have a well-established discipline of evaluation, just as we do of measurement or of statistics or experimental design’ (1996, p. 401). Scriven went on to suggest that what a discipline required included ‘a basic conceptual framework—a low-level theory...’, which he believed evaluation did in fact have, along with ‘four basic predicates: grading, ranking, scoring and apportioning’ (p. 401).

Despite Scriven’s exhortations, Kushner’s subsequent work indicates that the ‘paradigm wars’ are a reality for evaluation, and indeed argues against a postmodern desire to annihilate paradigmatic distinctions (2002). Such efforts fail to recognise paradigm wars as focussing on methods rather than values—for Kushner, what is critical for evaluators is that their work be based on well-understood values (ie methodology, in his terms), not that they prioritise the question of methods of evaluation (2002). It is possible, for some, however, to discern movements in both methodology and methods consistent with waves of education reform, shifting from a positivist paradigm to an interpretive perspective; from a behaviorist theory of learning to a constructivist one; from emphasising external incomes and factors to the concern for school procedures and pedagogical practices inside the classroom (Martinic, 2012). These developments may signal changes in values too, from a punitive and authoritarian value base to values that resonate with concepts of social justice and democratic participation, values that Kushner argues as a sound basis for evaluation (2002; 2005). The preceding discussion thus suggests a valid question to be whether ERO has a values base, and what its particular methodological (if not theoretical) approach may be.

Values and Theory in ERO

The most direct statement of values in the practices of ERO is contained in its *whakataukī* (proverb, or motto), namely: *Ko te tamaiti te pūtake o te kaupapa: The child - the heart of the matter* (n.d. a). Its stated aim is to improve the achievement of all students, by evaluating the quality of schooling (and early childhood services) in New Zealand (ERO, 2011b). It is committed to honouring the Treaty of Waitangi (the founding document of the New Zealand nation, which outlines the agreement between the Crown and Māori) in its work. Its handbook,

'Evaluation Indicators for School Reviews', specifically refers to the prioritisation of the interests of Māori and Pacific Island students in its reviews. This position is reinforced by the intention of ERO to ensure that schools are responding to student diversity, and raising the achievement of all students (2011b, pp. 4–5). Thus, ERO is not an agency that is established to deliver dispassionate or objective advice or comment on the delivery of programmes, but clearly to operate within a strictly limited remit as set out in legislation, and demanded by the government of the day. Nevertheless, ERO officers would see themselves as performing a task of significant democratic import (providing clear and current information to families seeking knowledge of their local or like schools of choice) in a manner that supports a social justice imperative (the prioritisation of Māori and Pacific Island student attainment).

Scriven (1996) laid claim to a conceptual framework as adequate evidence that an evaluation service had a theory of practice. In the case of ERO, this framework is referred to as the 'six dimensions of good practice' (2011b, p. 6), and is underpinned by its own accumulated experience of 'effective schooling', meta-analyses of effective pedagogical and leadership practice, and 'current evaluation theory' (p. 6). Evidently, therefore, its epistemological frame of reference is firmly situated in the discourse of school effectiveness, as further evidenced by the details of the dimensions of good practice: Student learning: engagement, progress and achievement; effective teaching; leading and managing the school; governing the school; safe and inclusive school culture; engaging parents, whānau and communities (p. 6). In its graphic depicting this conceptual framework, student learning is the large central core, around which the other five conceptual elements are gathered (p. 7).

The evaluation theory that has guided the work of ERO in the preceding two or three years is 'complementary review'. It is the practical application of this theoretical approach that ERO appears to have softened its previously strict approach to accountability. From the perspective of ERO, the school has a legitimate and valued role to play in the school review process, using the tool of self-review, which has been part of school practice for many years prior to the uptake by ERO of complementary review. Thus the results of a school's own internal review processes are intended to complement (or be complemented by) the outcomes of the external accountability review process conducted by ERO (p. 7). What is the value or purpose of complementarity? For Feinstein (2012), it is the most desirable state in which evaluation can find itself, as it is in this moment that lessons can be learned. This learning is only made possible when accountability evaluation provides an incentive by treating errors in practice as lessons to be learned (rather than as pretexts for punishment, presumably). In the context of complementary review, sound evaluation strategies seek to assign attribution, that is, they are strategies motivated to locate the causes of the successful achievement of outcomes. Feinstein argues that these strategies work best when guided by the criteria of relevance (ensuring a coherent link between action and intended outcomes), effectiveness (assessing the extent to which outcomes are achieved), and efficiency (assessing the cost of attaining the outcomes) (2012, pp. 106-108).

This shift towards complementarity may be seen as a counter to the heavy-handedness of the accountability regime associated with Public Choice Theory (PCT), itself a manifestation or earlier neoliberal policymaking. In this development, the state created quasi-market conditions to influence public-sector transactions. This was achieved in part by uncoupling services from their parent organisations, and in some cases (such as ERO), agents separated from the parent organisation (in this case the Ministry of Education) were contracted to apply rigorous accountability from policy-making to implementation within the parent organisation or amongst its other agencies (in this case, schools) (Olssen, Codd & O'Neill 2004; Olssen & Peters 2005; Peters 1999). Lehtonen (2005) has suggested that the certainties imposed by accountability regimes have been corrected, to some extent, by the uncertainties, open-endedness and plurality that he links to sustainability discourses (incidentally a feature too of ERO thinking, especially in relation to the longevity of good practice in schools). This correction has led to a call for revised ways of evaluating policies, for instance, in ways that provide a credible opportunity for those being evaluated to have a voice and to learn by the process. Nevertheless, this position is contested, as learning may be seen to trump accountability, the very purpose of evaluation (2005). Leaving aside for now questions of critique, which will be considered later, I wish to turn now to consider matters concerning the methods by which ERO conducts its complementary reviews.

The Education Review Office Method

What follows attempts to describe, without critical comment, highlighted aspects of the way in which ERO reviews schools, with some reference to the way it structures its reports, the role of qualitative and quantitative methods in ERO reviews, ending with the seemingly admirable principles and guidelines by which its reviews are designed.

The approach taken by Education Review Office generally follows a set of routines (ERO, 2011a, p. 8) that vary little—a school is alerted well in advance by letter that it is to be reviewed at some point over the next two school terms; the principal and key board of trustees members are invited to a general information session; the school is required to gather its samples of the data and information as evidence it is meeting its legislative requirements, including the crucial *Board Assurance Statement and Self-Audit Checklists*; the school is alerted nearer the time of the actual date of review; some two to three days may be set aside for the review (school size dependent); usually two or three officers will be assigned; a pre-review meeting between the lead reviewer and the board and key staff occurs, when the expectations of the reviewers is outlined and the review is scoped and designed around the following major question: ‘How effectively does this school’s curriculum promote student learning – engagement, progress and achievement?’ (ERO, 2011a, p. 4). In addition, data will be gathered in relation to specified topics of national importance, which rotate during each year. These ‘National Evaluation Topics’ (n.d. b) become the source of meta-analysis national reports published by Education Review Office, based on a sample of schools visited in a given year. During the school visit, officers interview the board and key staff, meet the whole staff (usually without the school leaders being present), interview community members, including a selected group of students, attend meetings as observers, observe lessons, and review large amounts of collected documentary evidence. Over these days, the officers will monitor findings and issues as they emerge, and discuss with one another this emergent meaning, sometimes seeking further clarification from school staff. By the end of the review period, this emergent understanding becomes a shared understanding with the school, and a draft report is compiled in answer to the major evaluation question. ‘Shared understanding’ in this context means that the key members of the school will ‘work constructively with the review teams to identify the implications for action, areas for development and review and to develop any recommendations or actions for compliance based on Education Review Office’s findings’ (p. 10).

The findings are reported to the school in a short (about 8 page) qualitative report that is structured under the headings of context, learning, curriculum, and sustainable performance (see <http://www.ero.govt.nz/Early-Childhood-School-Reports> for a sample of current reports). The report is presented to the school board in ‘unconfirmed’ form, and, after a period of time in which the board can make further comment (sometimes leading to textual changes to the report), it is confirmed, and made available on the ERO website. These reports are expressed in generally value-neutral terms and use factual language. Terms of approbation are controlled, and may be expressed using the adjectival forms ‘very well’, ‘well’, ‘good’ and ‘committed’, while adverbial forms commonly used include ‘positively’, ‘actively’, ‘effectively’ and ‘productively’. In regard to recommendations for improvement, common terms include, ‘could now’, ‘evaluate’, ‘should continue to’, ‘more’, ‘further’, ‘develop more’, ‘access (advice)’ and ‘include’.

Despite its data gathering including quantitative school reports, (usually analyses of student achievement data), the ERO reports to schools do not contain any quantitative analysis. In part, this may be understood in relation to the intended audience of these reports, which is not only the school staff, but also its board, made up usually of parents, and the wider community of families, as well as prospective families searching for a suitable school. In contrast, its national reports are substantial documents (see, as an example, 2012), and may include reference to sampling size and method, methodology and research evidence. As these reports are meta-analyses of a significant sample of schools reviewed in a given period, ERO makes statistical comparisons of general findings across the sample. The body of a typical report will include qualitative findings sourced from meetings and interviews in exemplar schools, and will engage in interpretive discussion, leading to recommendations for action to the education sector (including government). These national reports serve the significant function of delimiting ‘best practice’ in the specific areas of national interest.

The design and conduct of school reviews, apart from being underpinned by the 'six dimensions of good practice' conceptual framework, is governed by a set of principles and guidelines (2011 a, p. 7). These appear to support admirable principles of social justice (a focus on Māori and Pacific Island achievement), critically reflective practice embodied in forms of practitioner action research, and democratic qualities such as transparency, collaboration and participation. Finally, they reflect a commitment by ERO to have realistic expectations of school's, and to support them to use the evaluation process as a transformative opportunity. Despite this hopeful note, I wish to suggest, however, that there remain areas in ERO's approach requiring critical comment.

Critique

The following critique begins from the premise that the work of ERO remains fundamentally based on a low-trust accountability model that seriously undermines its newly-found democratic intent. Its epistemology is one founded on the discourse of school effectiveness and improvement, which has implications for the assumptions and methods guiding its work. Together, these flaws reflect complementarity for what it is—an exercise in subjectification to a regime of self-discipline and self-punishment.

Evaluation is not neutral (Kushner, 2002). This is the case, in part, because it implies an evident power differential (the evaluator has power over the one being evaluated). For some, evaluation is threatening (Feinstein, 2012; Thrupp, 1998). Thrupp's (1997) position was that ERO 'hunts down' failing or struggling schools, and suggested that it participates in the 'politics of blame' (1998). It may be suggested that Thrupp's critiques of Education Review Office are dated, and given what has been described above, ERO has come some way towards creating an experience for schools that is better balanced and more rewarding. Nonetheless, Thrupp concluded more recently (2008) that ERO engages in the 'politics of blame', which supports neoliberal marketisation discourse that regards 'failing' schools as the authors of their own misery rather than considering the negative effects of ideologically driven policy on schools and teachers.

While ERO's place and role in society owes much to the neoliberal, market-oriented reforms of the late 1980s and early 1990s, there is a shift, much as suggested by Martinic (2012), in evaluation strategies and purposes aligned to increasing sophistication in the policies of neoliberal governments. Whereas a key role for ERO may have been (and still continues to be) providing transparent information to support parental choice in the education quasi-market, the current strategic focus on the enhancement of scholastic achievement by underperforming groups in society is echoed by ERO, which sees the purpose of its reviews being 'to contribute to improved student achievement... and [to] give schools external evaluative information to support improvement' (2011a, p. 1).

As the task of evaluators contracted to government is to evaluate the implementation of its policies, such evaluation must thus be on the government's terms, which not only precludes dissent (Kushner, 2002) but also reinforces the power differential already referred to, regardless of the discourses evaluators, such as ERO, engage in. The accountability demands of stakeholders such as government, means that evaluators find themselves as protectors of the status quo (Stake, 2001). It is in attempting to meet the remit of its employer (the state) that ERO engages in methods that overshadow any desirable democratic intent. These methods include the stipulated accounting to ERO by school boards providing assurance of their compliance with various acts of legislation and policy imperatives, requiring the completion of a 20 page checklist that is supported by a 50 page guide (2013). The relevant *Board Assurance Statement and Self-Audit Checklist* reminds schools that '[t]he more assurance a board can provide to ERO that it is meeting its legal requirements and taking appropriate action to remedy areas of non-compliance, the greater the emphasis can be on other areas of review' (2013, Instruction Page). Further, the board is encouraged 'to note areas where you are aware that you are not meeting legal requirements and to advise any action you are taking' (p. 2). The board chairperson and principal sign a certificate of compliance, attesting to the board's legal compliance with the relevant legislation, and necessary remedial action. There is little (apart from tragedies) in the life of New Zealand schools that focuses the mind like the knowledge of an impending ERO visit.

Wrigley suggested that the terms *effectiveness* and *improvement* are so ideologically loaded that disagreement is a 'sheer impossibility... you could no more wish to be 'ineffective' or reject the call to 'improve' a school, than you could disagree with personal hygiene or kindness to animals' (2004, p. 36). While school effectiveness (SE) is a quantitative attempt to measure a school's performance against certain criteria, school improvement (SI) is a qualitative study of the development of leadership, management, school culture and *teaching* to establish how these contribute to student attainment (2004). ERO is committed to this discourse by its support of continuous improvement and the priority it gives to validating the effectiveness of school curriculum programmes (see, for example, 2011a, p. 17). Schooling is rendered as a technical-rationalist 'enterprise' by this approach, thus removing any moral aspects from debates about schooling. However, as education is a moral enterprise, educators ought to raise questions about educational desirability rather than effectiveness (Biesta, 2007).

Elliott (2009, p. 175) drew attention to the decontextualised character of the SE/SI by pointing out that teacher effectiveness research abstracts co-relational variables from particular contexts, which provides teachers no certainty within their unique contexts, a point also made by Biesta (2007). Allied to SE/SI discourse is the mantra of 'evidence-led teaching' or 'data-led practice', to which ERO is also committed. Examples include the questions asked in relation to its evaluation indicators of 'effective teaching' (2011b, p. 20) and 'leading and managing the school' (p. 27). Biesta (2007) notes this focus on evidence has come into fields like education from the medical sciences, but contests the validity of that move. One of his reasons (apart from the lack of likeness between patients and students), is that research (read evaluation, in this context) and the inquiries of others can only tell us what worked in other situations, but those findings cannot be a rule to prescribe all future transactions. Biesta would therefore challenge the value of the ERO national reports, and would suggest that school review reports are of historical value only. Furthermore, however, he argued that the notion of 'evidence-led' education and effectiveness has a narrowing effect, with a concentration on the determination of causality. The question of effectiveness is an instrumental question, based on the idea that a professional act in every situation can bring about an effect (2007). Indeed, to see teachers' actions as the cause of learning is deeply behaviouristic, and implies faith in a mechanistic input-output model. For Martinic, the underlying theory of action of such discourses is that information can produce changes in practices. (2012).

To ensure this mechanistic narrowness of purpose, and presumably to assure replication of review processes, ERO functions with a set of clearly articulated evaluation indicators (based on its 'six dimensions of good practice') that are further sub-divided into themes, each one providing review officers with a number of question prompts, exemplar indicators, and suggestions for relevant evidence (2011 b, pp. 15–50). The danger inherent in evaluation indicators is that they obscure what is significant and interesting in a case (Stake, 2001), while Kushner, who sees evaluation ideally serving to highlight the best in the public service, has suggested: 'By asking evaluation to focus so relentlessly on outcomes we have too few complex accounts of the quality of public works' (2011, p. 312). Indicators predetermine ends in deterministic language, something Biesta abhors—ends and problems should only be expressed as hypotheticals (2007).

Finally, it must be asked if the concept of complementarity is capable of generating democratic dialogue and process between a school community and ERO? Kushner, who endeavours to promote a positive view of evaluation as a discipline and profession, indicated that evaluation has an obligation to neutralise power differences in the case being studied (2002). Furthermore, in relation to review design, he has argued against strong designs, favouring instead emergent designs, which are flexible and allow meaning to emerge (2005). While this appears to be partly consistent with ERO practice, much of its design looks rather like Kushner's strong (anti-democratic) design, which prepares judgment criteria in advance, decides on samples beforehand, and seeks to ensure coherence beforehand. 'The difference [between the two design types] is profound, as [emergent design] requires the evaluator to share (if not entirely cede) intellectual control over an evaluation with his or her respondents' (p. 580).

Insights offered by Foucault provide additional lines of critique and understanding of complementarity. The sharing of control hinted at by Kushner is not motivated by ERO's largesse; the act of complementarity is made possible because the unique role ERO has to play in New Zealand society empowers it within the range of services it reviews. Correspondingly, however, as power is only possible over a subject who has freedom, or at

least a range of choices (Foucault, 1994b), complementarity is a game that requires two to play, and by participating, boards and schools have the option to engage in some transformative acts. While Foucault's reasoning that there exists an implicit agreement on what constitutes right behaviour in these contexts (Barker, 1998) may seem to imply acquiescence by those without power, in fact, argued Heller (1996), resistance is equally possible. Complementarity at least opens up some creative possibilities for more balanced engagement between ERO and the school and its board.

More problematically is the subjectification of the board and school to a regime of self-discipline and self-punishment provoked by the 'internal' aspect of complementarity. Feinstein (2012) has suggested that outcome indicators permit the notion of complementary learning as it is possible to see what one has achieved—or failed to achieve—by evaluating one's progress against the indicators. Foucault's historical analysis suggested to him that care of the self was a primary objective in Antiquity, with emphasis being placed on spending time with oneself to better understand oneself (1994 a). This, it may be implied, indicates a form of self-management (so that one does not have to be managed by others) and is what a school and its board is required to exercise, to successfully learn by its errors (and achievements). In monastic Christianity, Foucault understood obedience and contemplation to be important principles of self-management. The obedience of the monk to his master was complete and lifelong, demanding complete self-sacrifice of the individual to the master. This obedience and self-sacrifice is a technology of the self—indeed, a technology of self-examination. (1994a). Self-review in New Zealand schools, I argue, are precisely such a technology. Furthermore, the regime of self-review ensures, at no cost to the state (or ERO), on-going self-surveillance and self-discipline by the school.

Conclusion

In this article, I have argued that the New Zealand Education Review Office, established under the auspices of a neoliberal regime, despite shedding some of the conceptual and procedural baggage that implies, remains committed to working with a low trust model. This article has attempted to locate the work of ERO in the framework of evaluation theory, and considered the theoretical assumptions of ERO itself. While ERO is motivated by a social justice imperative, and has begun to engage with schools on the basis of complementarity, its methodology reflects a commitment to the discourses of school effectiveness and improvement, an approach, it was argued, that ignores the possibility of education as a moral endeavour. In an extended critique, these points were considered in greater detail, and with reference to some Foucauldian ideas, the concept of complementarity was shown to have both the potential for the creative and transformative use of power by boards, yet to disguise a technology of incessant self-examination. That tension appears to now be a reality that boards and schools cannot escape.

References

- Barker, P. (1998). *Michel Foucault: An introduction*. Edinburgh, United Kingdom: Edinburgh University Press.
- Benade, L. (2009, 18 Dec). Worrying signs of teacher-bashing in latest ERO report. The New Zealand Herald. http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=10616219
- Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory* 57(1), 1–22.
- Codd, J. (2005). Teachers as 'managed professionals' in the global education industry: the New Zealand experience. *Educational Review*, 57(2), 193–206.
- Dale, R. (1999). Specifying globalization effects on national policy: a focus on the mechanisms. *Journal of Education Policy*, 14(1), 1–17. DOI:10.1080/026809399286468

- Dale, R. and Robertson, S.L. (2002). The varying effects of regional organizations as subjects of globalization of education. *Comparative Education Review*, 46 (1), 10-36. DOI: 10.1086/324052
- Education Review Office. (ERO, 2011a). Framework for school reviews. Retrieved 24/10/2013 from <http://www.ero.govt.nz/Review-Process/For-Schools-and-Kura-Kaupapa-Maori>
- Education Review Office. (ERO, 2011b). Evaluation indicators for school reviews. Retrieved 5/7/2013 from <http://www.ero.govt.nz/Review-Process/Frameworks-and-Evaluation-Indicators-for-ERO-Reviews>
- Education Review Office. (ERO, 2012). Teaching as inquiry: Responding to learners. Retrieved 5/7/2013, from [http://www.ero.govt.nz/National-Reports/\(year\)/2012](http://www.ero.govt.nz/National-Reports/(year)/2012)
- Education Review Office. (ERO, 2013). Review Documentation for Schools. Retrieved 25/10/2013, from <http://www.ero.govt.nz/Review-Process/For-Schools-and-Kura-Kaupapa-Maori/Review-Documentation-for-Schools>
- Education Review Office. (ERO, n.d. a). About Us. Retrieved 22/10/2013, from <http://www.ero.govt.nz/About-Us>
- Education Review Office. (ERO, n.d. b). National Evaluation Topics. Retrieved 25/10/2013, from <http://www.ero.govt.nz/Review-Process/National-Evaluation-Topics>
- Education Review Office. (ERO, n.d. c). What ERO Does. Retrieved 21/10/2013, from <http://www.ero.govt.nz/About-Us/What-ERO-Does>
- Elliott, J. (2009). Research-based teaching. In S. Gewirtz, P. Mahony & I. Hextall (Eds.), *Changing teacher professionalism: International trends, challenges and ways forward* (pp. 170-183). London/New York: Routledge Taylor and Francis Group.
- Feinstein, O. N. (2012). Evaluation as a learning tool. In S. Kushner & E. Rotondo (Eds.), *Evaluation voices from Latin America. New Directions for Evaluation*, 134, (pp.103–112). DOI: 10.1002/ev.20022
- Foucault, M. (1994a). Technologies of the self. In P.Rabinow (Ed.). *Ethics: Subjectivity and truth: Essential works of Foucault. Volume One.* (pp. 223–252). Translated by R. Hurley and others. London, England: Allen Lane The Penguin Press.
- Foucault, M. (1994b). The subject and power. In J.D. Faubion (Ed.). *Power: Essential works of Foucault. Volume Three.* (pp. 326–348). Translated by R. Hurley and others. London, England: Allen Lane The Penguin Press.
- Gordon, L. (1997). ‘Tomorrow’s Schools’ today: school choice and the education quasi-market. In M. Olssen & K. Morris Matthews (Eds.), *Education policy in New Zealand: the 1990s and beyond* Palmerston North: Dunmore Press.
- Heller, K.J. (1996). Power, subjectification and resistance in Foucault. *SubStance*, 25(1), 78-110. Stable URL: <http://www.jstor.org/stable/3685230>
- Kushner, S. (2002). I’ll take mine neat: Multiple methods but a single methodology. *Evaluation*, 8(2), 249–258.
- Kushner, S. (2005). Democratic theorizing from noun to participle. *American Journal of Evaluation*, 26 (4), 579-581 DOI: 10.1177/1098214005281357

- Kushner, S. (2011). A return to quality. *Evaluation*, 17(3), 309–312. DOI: 10.1177/1356389011410215
- Lehtonen, M. (2005). OECD Environmental Performance Review Programme: Accountability (f)or learning? *Evaluation*, 11(2), 169–188. DOI: 10.1177/1356389005055536
- Martinic, S. (2012). Emergent evaluation and educational reforms in Latin America. In S. Kushner & E. Rotondo (Eds.), *Evaluation voices from Latin America. New Directions for Evaluation*, 134, (pp. 17–28).
- McKenzie, D. (1997). The cult of efficiency and miseducation: Issues of assessment in New Zealand schools. In M. Olssen & K. Morris Matthews (Eds.), *Education policy in New Zealand: the 1990s and beyond* (pp. 47–64). Palmerston North: Dunmore Press.
- New Zealand Government. (2009). Education Act 1989 No 80. Retrieved 21/10/2013, from <http://www.legislation.govt.nz/act/public/1989/0080/latest/DLM187387.html>
- Olssen, M., Codd, J. & O'Neill, A-M. (2004). *Education policy: Globalisation, citizenship and democracy*. London, United Kingdom: SAGE Publications.
- Olssen, M., & Morris Matthews, K. (1997). Introduction. In M. Olssen & K. Morris Matthews (Eds.), *Education policy in New Zealand: the 1990s and beyond* (pp. 7–46). Palmerston North: Dunmore Press.
- Olssen, M., & Peters, M. (2005). Neoliberalism, higher education and the knowledge economy: From the free market to knowledge capitalism. *Journal of Education Policy*, 20(3), 313–345.
- Peters, M. (1999). Neoliberalism. In M. Peters, P. Ghiraldelli, B. Žarnić, & A. Gibbons (Eds). *A Encyclopaedia of Philosophy of Education*. <http://www.ffst.hr/ENCYCLOPAEDIA/doku.php?id=neoliberalism>.
- Scriven, M. (1996). The theory behind practical evaluation. *Evaluation* 2(4), 393–404. doi:10.1177/135638909600200403
- Snook, I. (2003). *The ethical teacher*. Palmerston North: Dunmore Press.
- Stake, R.E. (2001). A problematic heading. *American Journal of Evaluation*, 22(3), 349–354.
- Thrupp, M. (1997). Shaping a crisis: the Education Review Office and south Auckland schools. In M. Olssen & K. Morris Matthews (Eds.), *Education policy in New Zealand: the 1990s and beyond* (pp. 145–161). Palmerston North: Dunmore Press.
- Thrupp, M. (1998). Exploring the politics of blame: school inspection and its contestation in New Zealand and England. *Comparative Education*, 34(2), 195 - 210.
- Thrupp, M. (2008). *Secondary teaching, social contexts and the lingering politics of blame*. Paper presented at the New Zealand Post Primary Teachers' Association professional conference, 'Secondary teaching on the move', Waipuna Hotel and Conference Centre, Auckland.
- Wrigley, T. (2004). School effectiveness and improvement: where is the grand narrative? In J. Satterthwaite, E. Atkinson & W. Martin (Eds.), *Educational counter-cultures: confrontations, images, vision*. Stoke on Trent, United Kingdom: Trentham Books.