

Evidence of Teaching Practice in an Age of Accountability: When What Can be Counted isn't All That Counts

Nicole Mockler, Sydney School of Education and Social Work, University of Sydney

nicole.mockler@sydney.edu.au

ORCID: 0000-0002-0617-3506

Twitter: @nicolemockler

Meghan Stacey, School of Education, University of New South Wales

m.stacey@unsw.edu.au

ORCID: 0000-0003-2192-9030

Twitter: @meghanstacey

Abstract

Cultures of performative accountability in education have been on the rise globally since the 1980s. Accordingly, teachers have increasingly been encouraged to understand their work in relation to particular forms of 'evidence'. All evidence, however, is not regarded as equal, and sources of evidence privileged within cultures of performative accountability are typically narrowly rendered, apparently 'objective', externally generated and largely quantified in form. Recent research from the United States, Australia and England has suggested that such cultures may be re-shaping teacher practice and identity. Performative accountability, however, is not the only form of accountability, and this paper explores the tensions between performative and 'intelligent' accountability in teachers' work and lives, reporting on research investigating Australian teachers' understanding of and engagement with data and research as well as evidence of their work more broadly. The paper contributes to the literature on the datafication of teaching and debates regarding the nature and role of evidence in teaching work. Ultimately, we argue that spaces for rich understandings of evidence, present but not dominant in contemporary education policy, must be protected and extended in order for the current focus on evidence to generatively – and intelligently – reflect and shape teachers' work.

Accountability in education, like 'quality', can be difficult to argue against. Nobody, it seems – from politicians to teachers themselves – is in favour of the 'unaccountable teacher'. Not all forms of accountability, however, are equally purposeful or useful, and this paper, working with the voices of practising teachers in Australia gathered via interviews and questionnaires, troubles contemporary notions of accountability in education and the usefulness or otherwise of dominant accountability approaches to teachers' work. We argue that what 'counts' for teachers as good evidence of practice is not only that which can be counted. Far from eschewing accountability itself, the teachers in our study pointed to the importance of rich, nuanced sources of evidence more characteristic of alternative, 'intelligent' (O'Neill, 2002, 2013) forms of accountability. Consequently, we argue that spaces for rich understandings of evidence, present but not dominant in contemporary education policy, should be protected and extended in order for the current focus on evidence to generatively – and intelligently – reflect and shape teachers' work.

In making this argument, we consider a broad spectrum of 'evidence' in teaching that ranges from narrowly conceived ideas about numeric 'data', to considerations of evidence of teaching practice generated as part of practitioner engagement 'in' and/or 'with' academic research (see Prendergast & Rickinson, 2019). Our aim is to engage with a range of ways of seeing, understanding and evaluating

teaching work, and in doing so to consider the different ways in which particular forms of evidence are mobilised in the pursuit of different accountability purposes.

The paper begins with a discussion of the role of evidence in education, including its potential implications for teachers and its relationship to different forms of accountability. We then introduce our study and explore the teacher participants' orientations to research and data, and perspectives on what constitutes good evidence of practice, using the lens of different forms of accountability as an analytical tool.

What 'counts' as evidence in education?

Across the globe, particular forms of 'evidence' in education are taking on increasing prominence (Helgetun & Menter, 2020; Malone & Hogan, 2020; Torrance, 2014). National large-scale assessments are on the rise internationally and are increasingly used for monitoring and evaluation purposes (Holloway et al., 2017; Verger et al., 2019). Elsewhere, teachers are encouraged to generate and use 'gold standard' evidence via randomised controlled trials (Goldacre, 2013). As Ladwig (2018) reminds us however, 'evidence' is never neutral. The idea that 'evidence isn't evidence until we socially recognize it as evidence' (McKnight & Morgan, 2019, p. 655) suggests that 'what counts' is constructed rather than pre-existing and self-evident. In education, large-scale quantitative data are often more valued than qualitative, small-scale data (Welch, 2015). In this new, datafied landscape counting becomes a primary way of understanding and evaluating the self, part of both individual and wider social life and governance (Rose, 1991). Yet as Power (2004) reminds us, 'much has to be done to render diverse phenomena countable quanta in the first place, namely an abstraction from many specific qualities by establishing categories of similarity' (p. 767); such processes can serve to obfuscate nuance and complexity, and are therefore worthy of attention.

Applying such attention to research regarding teachers' practice – where it should come from and how it should be used – reveals similar trends. Cain and Allen (2017) find that much educational research remains 'invisible' to teachers as 'impact' has largely been understood as the capacity for research to shape policy and policymakers' decisions. Biesta (2007) has contended that calls for narrowly conceived forms of evidence-based practice in education entail two interrelated claims:

that *educational research* should not be left to educational researchers but should be subject to centralized agenda-setting...so that it can become more practically relevant.

and reciprocally:

that *educational practice* should not be left to the opinions of educators but that their work should be based upon research evidence. (p.2)

Particular forms of 'research', and indicators of teachers' engagement with such research, can thereby become another external mechanism through which teachers' work is judged, alongside the data produced in teachers' classrooms. In both, the process appears to be about finding ways of 'bracketing out' teacher professional judgement, said to be subjective, and thus inferior to more 'objective' ways of knowing and assessing.

Leat et al. (2015, p. 274) argue that in recent years, teachers' engagement with and in research has experienced a 'strong gravitational pull towards school effectiveness approaches'. This contrasts with the ideals of 'critical autonomy' embedded in previous views of teachers' engagement with research, and, as Godfrey (2017) has observed, stands in contrast to the more 'empowering vision' of the relationship between educational research and teachers' work embedded in recent white papers developed in the UK (Furlong et al., 2014) and Australia (White et al., 2018).

Mockler and Groundwater-Smith (2018) write of the 'alluring simplicity' of such declarative and prescriptive approaches, but argue that:

Eliminating the human dimension of education is not the answer. A more generative response would be to recognise that within this complexity, a teaching profession that understands what constitutes good (but varied) evidence of learning, and how to engage in true evidence-informed practice, building complex pictures of student learning to inform their judgement and decision making, is critical. (p. 49)

In the context of test-based or ‘performative’ (Lyotard, 1984) accountabilities, however, teachers risk becoming subject to regimes of narrowly-rendered forms of evidence, and their work open to reconstitution and reconfiguration at the hands of these regimes, as recent research has shown (Hardy & Lewis, 2017; Holloway & Brass, 2018).

Teachers as subject to regimes of evidence

Much of the research relating to teachers’ subjection to regimes of evidence has focused on a push for teachers to engage with and be assessed through particular, numeric forms of data relating to their classroom practice and the students they teach. The literature suggests a range of ‘responses’ by teachers to such shifts (Holloway et al., 2017). Ball (2003) describes a phenomenon of ‘values schizophrenia’, where ‘there is a potential ‘splitting’ between the teacher’s own judgements about ‘good practice’ ... and the rigours of performance’ (p. 221). More recently in the context of English primary schooling, Braun and Maguire (2018) have written of teachers ‘doing without believing’, experiencing ‘a sense of deep unease’ (p. 8) attendant to their required compliance with assessment-related policy demands.

Yet Braun and Maguire (2018) also argue that binaries between being policy focused on the one hand, or holding onto ‘child-centredness’ on the other, can be too simplistic; of course, teachers can and frequently do believe in some merits of both simultaneously. In Australia, Hardy and Lewis (2017) identify what they describe as the ‘doublethink’ of data. Teachers in their study held ‘conflicted positions’, seeing the data they delivered as ‘worthless yet important, unnecessary yet indispensable, distracting but beneficial’; these teachers, it seemed, ‘could both deny and deify data, and deliver the data even as they questioned the benefits of doing so’ (p. 682). Seen in this way, teachers may be understood as moving between the positions of those upon whom policy regimes are unhappily enforced and experienced, and those who may take up and concur with such regimes. Some have argued that the latter is becoming an increasingly prominent way in which teachers understand themselves and their work. In a publication based on US data, Holloway and Brass (2018) compare two studies, conducted nearly a decade apart, identifying an ‘onto-epistemological’ shift towards new teacher subjects for whom data and related forms of performative accountability have become ‘constitutive’ (p. 378) of their professional identities.

Accountability and education

Producing, presenting and defining oneself in relation to particular forms of ‘evidence’ of practice, whether through data related to one’s students, association with pre-existing research proclaiming the ‘effect size’ of particular teaching practices (see e.g. Hattie, 2012), or even engaging in richer forms of practitioner research, is about presenting one’s work in particular ways and thereby, making oneself accountable. Just as there are different forms of evidence, however, there are different kinds of accountability. Much has been written about the various forms of accountability that have come to dominate education globally (see e.g. Hardy et al., 2019; Lingard et al., 2017; Mockler & Groundwater-Smith, 2018). O’Neill’s (2002) notion of ‘intelligent accountability’, initially developed in the context of her *A Question of Trust* Reith Lectures, has been a widely-used resource in this literature, giving rise to discussion of ‘alternative’ (Lingard et al., 2017) and ‘authentic’ (Hardy et al., 2019) counter-perspectives to dominant forms of accountability.

Lingard et al. (2017) argue that performative accountability, which they characterise alongside ‘consumer accountability’, ‘contract accountability’ and ‘corporate accountability’ as a form of ‘test-

based accountability' is currently the 'dominant mode of accountability in schooling' (p. 10). At the heart of performative accountability is the drive toward efficiency and effectiveness, and the desire to maximise the relationship between inputs and outputs (Lyotard, 1984). Furthermore, there exists a complex relationship between performative accountability, technologies of audit, trust, risk, evidence, and judgement. Power (1994, 1997, 2007) has written extensively about some of these relationships, noting that 'audits are needed when accountability can no longer be sustained by informal relations of trust alone but must be formalised, made visible and subject to independent validation' (1994, pp. 9-10). Elsewhere he writes that 'internal control systems have become central to a 'regulatory epistemology' in which demands for trust create corresponding demands for evidence' (Power, 2007, p. 39). O'Neill herself, writing of performative accountability, suggested the following relationship between accountability and trust:

Perhaps the present revolution in accountability will make us all trustworthier. Perhaps we shall be trusted once again. But I think that this is a vain hope – not because accountability is undesirable or unnecessary, but because currently fashionable methods of accountability damage rather than repair trust. If we want greater accountability without damaging professional performance we need intelligent accountability...Intelligent accountability, I suspect, requires more attention to good governance and fewer fantasies about total control. (O'Neill, 2002, pp. 57-58)

Within regimes of performative accountability, evidence operates as a technology of audit, with an adversarial (Groundwater-Smith & Mockler, 2009), proof-driven conceptualisation of evidence at work. As such, evidence functions to 'give account', shifting trust away from trust in people to trust in processes of accountability themselves, and positioning professional judgement as 'subjective' and thus untrustworthy. Furthermore, conceptualising evidence in this way reaches toward a 'best practice' or 'what works' approach, where the markers of good practice are externally generated and evidence is used to identify how far teachers 'measure up'. Within an environment governed by more intelligent forms of accountability, however, good evidence is more expansively construed (Talbot, 2016), and is put to more forensic (Groundwater-Smith & Mockler, 2009) use, valued for what it can illuminate and problematise about practice. Within such a conceptualisation professional judgement is a critical tool, worthy of trust and able to be developed and honed. The use of evidence in this way is part of an ongoing discernment on the part of teachers of 'what counts' with respect to their practice.

Against this backdrop, we set out to explore teachers' engagement with educational data and educational research at a time when, internationally, growing importance was being accorded to narrow conceptualisations of evidence while at the same time performative accountability regimes for teachers were on the rise. How teachers understand their own work in relation to these regimes may reveal more productive ways forward in navigating new terrain of teacher evaluation. In the data we explore below, we show how teachers, in their discussion of evidence and what constitutes good evidence of practice, demonstrate the possibilities of a more 'intelligent accountability', as well as the kinds of constraints that can be associated with that which is more performative.

The teachers, educational data and evidence-based practice project

This article draws on data from across two phases of a research project known as *Teachers, Educational Data and Evidence-based Practice*, approved by the University of Sydney Human Research Ethics Committee in 2018. We were interested in exploring how Australian teachers understand and engage with educational research (as a broad term inclusive of both research by academics and practitioner research), as well as educational data, which in the context of current policy and research we expected may be understood as a narrower term in relation to quantifiable forms of information. The three overarching research questions for the project were: (1) How do teachers understand and engage with educational research and educational data?; (2) How do teachers understand and collect evidence of their practice?; and (3) What are teachers' professional learning

and development needs in relation to research and data literacy? In this article, we consider evidence in response to the first two of these three questions, drawing from interview and questionnaire data. We paint a picture of teachers who feel compelled to engage with and use different forms of educational research and data, and understand these sources in a range of ways.

Methods

The first phase of the research consisted of 21 interviews with practising teachers recruited via social media, specifically Twitter and teachers' Facebook groups. The interviews were semi-structured (Merriam, 2009) and took place either face to face or online, depending on participant preference. Each author conducted roughly half of the interviews, with analysis involving coding for themes, using both inductive and deductive coding (Miles et al., 2013). For deductive coding, a start list was derived from a comprehensive review of the literature relating to teachers' engagement with educational data and research, and these codes were expanded inductively throughout the analysis of interview data. In accordance with our ethics protocol, interview participants were asked to nominate a pseudonym by which they would be known in the reporting of the research. Interview data were used to shape the second phase of the research, which involved an online questionnaire. The questionnaire was designed to test the findings from the interview data with a broader group of teachers, and incorporated both questions developed by the researchers and, with permission, a number of items were adapted from the *Questionnaire on the Utilization of Research-based Information* (Lysenko et al., 2007). Feedback on the draft questionnaire was sought and gained from a small number of subject-matter experts, both researchers and practitioners, with the practitioners piloting the survey independently and providing specific feedback on the instrument as it would likely be received by teachers. On the basis of this the instrument was amended and then sent directly to those who had previously expressed interest, and shared broadly on social media in late 2018, garnering 524 responses. The questionnaire contained questions designed to elicit both qualitative and quantitative data. Qualitative data were analysed using the same approach as interview data, while quantitative data were analysed using both descriptive statistics, the focus of some of the reporting in this paper, and confirmatory factor analysis and structural equation modelling, to be reported in a subsequent paper.

Participants: background and orientation to research and data

The teacher workforce in Australia has an average age of 44-52 years and features a high proportion of female employees (80.9% primary and 58.4% secondary), according to the most recent *Staff in Australia's Schools* study (McKenzie et al., 2014). Teachers work across both public and private settings, which respectively enrol 65.7% and 34.3% of students (Australian Bureau of Statistics [ABS], 2020). Our study was open to all teachers who volunteered to participate. In both the interview and questionnaire phases of this study, the majority of participants (57% and 66%, respectively) were located in New South Wales, the most populous Australian state. A further 33% of interviewees and 25% of survey respondents were located in Victoria, Queensland and the Australian Capital Territory, while the remaining 10% and 9%, respectively, were drawn from South Australia, Western Australia, Tasmania and the Northern Territory. 71% of interviewees and 61% of questionnaire respondents currently worked in Government schools, and participants ranged in experience from 1 year to more than 25 years of teaching, holding a range of roles, including classroom teacher, middle leader, and deputy principal. While we initially anticipated that there might have been significant differences between the perspectives of respondents from different demographic groups (for example, public school teachers as compared to those who teach in non-Government schools, where different accountability requirements and expectations regarding data use might be in play), there was only limited variation in the data across these different groups – perhaps because of the common interest in and focus on research and data across the participant group as a whole.

Perhaps unsurprisingly, participants in the interview phase of the study largely declared themselves to be well-disposed to working with data and educational research. One participant, for example,

described herself as a ‘teacher nerd’ because of her interest in unpacking assessment data for herself and her colleagues, another as ‘a bit of a data nerd’ for the same reason. Similarly, a third participant described herself as a ‘data queen’, who enjoyed engaging with and making sense of data generated via the National Assessment Program – Literacy and Numeracy (NAPLAN)¹, for example.

Questionnaire respondents were asked to indicate how far they identified as a ‘data enthusiast’ and ‘data sceptic’, and a ‘research enthusiast’ and ‘research sceptic’. Tables 1 and 2 present participants’ responses to these questions.

		Data Enthusiast			Total
		Agree	Neutral	Disagree	
Data Sceptic	Agree	12	5	11	27
	Neutral	10	9	4	23
	Disagree	41	5	4	49
Total		63	18	19	

Table 1: Data enthusiast/sceptic

		Research Enthusiast			Total
		Agree	Neutral	Disagree	
Research Sceptic	Agree	8	1	3	12
	Neutral	9	9	3	21
	Disagree	59	5	4	67
Total		76	15	9	

Table 2: Research enthusiast/sceptic (%)

Participants were generally more well disposed towards research than data, with 63% declaring themselves to be data enthusiasts and 76% research enthusiasts. Conversely, over a quarter of the participants (27%) identified as data sceptics, while only 12% identified as research sceptics. While only 9% of the participants identified as non-research enthusiasts, 19% disagreed that they were data enthusiasts. This may reflect particular connotations surrounding the term ‘data’ which position it as more reified and performative, as opposed to research, which has been shown to be valued by teachers (if not regularly used – see Prendergast & Rickinson, 2019), and which may still be perceived as broader and primarily located in generally trusted spaces of academia. Identification as enthusiast and sceptic was not presented to participants as a continuum or a binary, however, it was recognised that it is possible to hold both positions at once (recalling, e.g., Ball, 2003; Braun & Maguire, 2018; Hardy & Lewis, 2017), and to this end a small group identified as both research enthusiast and sceptic (8%) and data enthusiast and sceptic (12%). Additionally, 4% of participants declared that they were neither research enthusiast nor sceptic, and 4% also declared themselves to be neither data enthusiast nor sceptic.

Survey participants were asked how important it was, in their view, for teachers to stay up to date on research about the subject/s they teach and about good classroom practice; to conduct small-scale inquiry into their practice; to deliberately gather evidence about their own classroom practice; and to use existing data within their school in their planning and teaching. Figure 1 below summarises responses to these questions, highlighting that a large majority of teachers (between 89% and 97%) within the sample indicated that each of these was either extremely, very or quite important for teachers.

¹ NAPLAN is the national standardised testing program in Australia. All school students in Years 3, 5, 7 and 9 sit for NAPLAN tests in May each year, and the results are disseminated to schools, parents and the general public via the Commonwealth Government’s MySchool website.

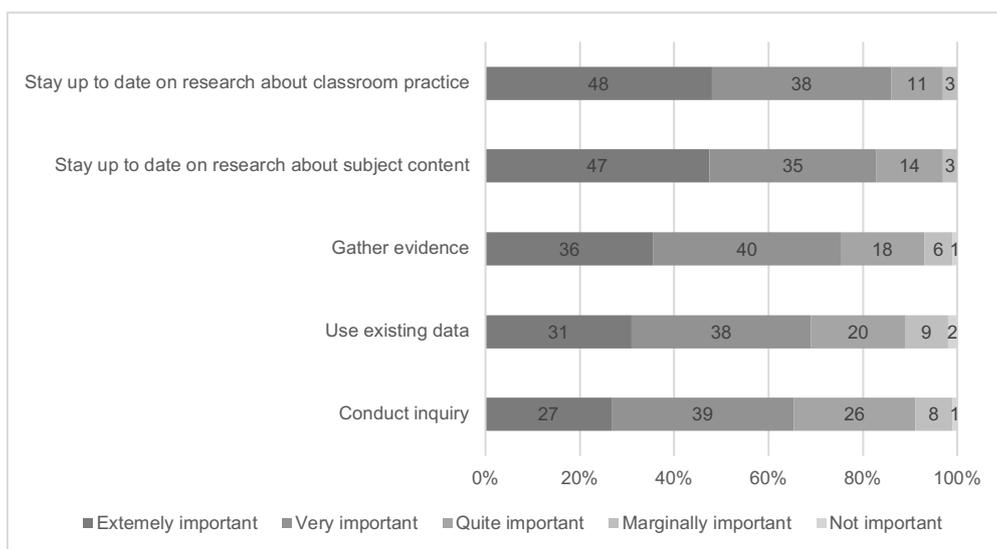


Figure 1: It is important for teachers to...

Furthermore, at least 50% reported that they were skilled at evaluating research, translating it into practice, interpreting qualitative data and collecting data from their own students about their learning, as highlighted in Figure 2 below.

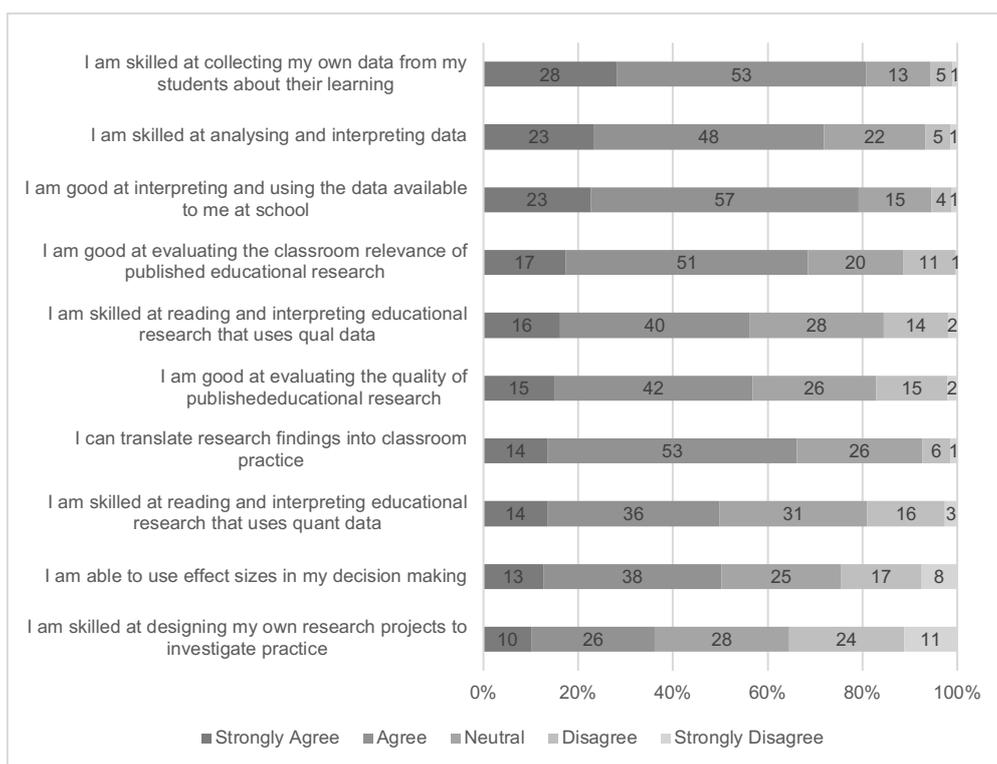


Figure 2: Teachers' reports of their skill levels

These results suggest that a particularly engaged sample was drawn in this study, likely a reflection of both the sampling technique used, garnering particular communities of teachers active in social media, and the phenomenon whereby participants with a particular perspective on a research topic will be more drawn to engage with the project. Qualitative data indicate nuance here, with some participants identifying their own orientation to be in contrast to that of the broader teaching population. This was the second most prevalent theme in the final open-ended survey question, raised in 13 out of 122 responses. In the words of one respondent, 'I see the importance in educational research, I see the importance of data, and I kind of wish that all people did in a way, and that's the sad bit of it all',

suggesting that they felt their views were not broadly held within the profession. This also resonated with our interview data.

Given the shape of the participant group, their orientation to the use of research and data to inform decision making and their reported expertise in doing so, our expectation was that their assessment of valid and reliable evidence of practice would be reasonably consistent with the kinds of evidence valued under regimes of performative accountability. Their responses, however, across both interview and open-ended survey responses, suggested a more nuanced and complex understanding of evidence, and it is to an exploration of these responses that we now turn.

‘Good’ evidence of teaching practice

Both in interviews and in the online questionnaire, participants were asked an open-ended question designed to elicit their views on what constitutes good evidence of their practice. This question was answered by all 21 interviewees and 265 of the 524 survey respondents (i.e. 51%). Additionally, in the online questionnaire, participants were asked how far they believed a range of different sources of evidence, generated both out of interview data and our broader reading of the field, to constitute valid and reliable evidence of teaching practice. Figure 3 below summarises their responses to this question.

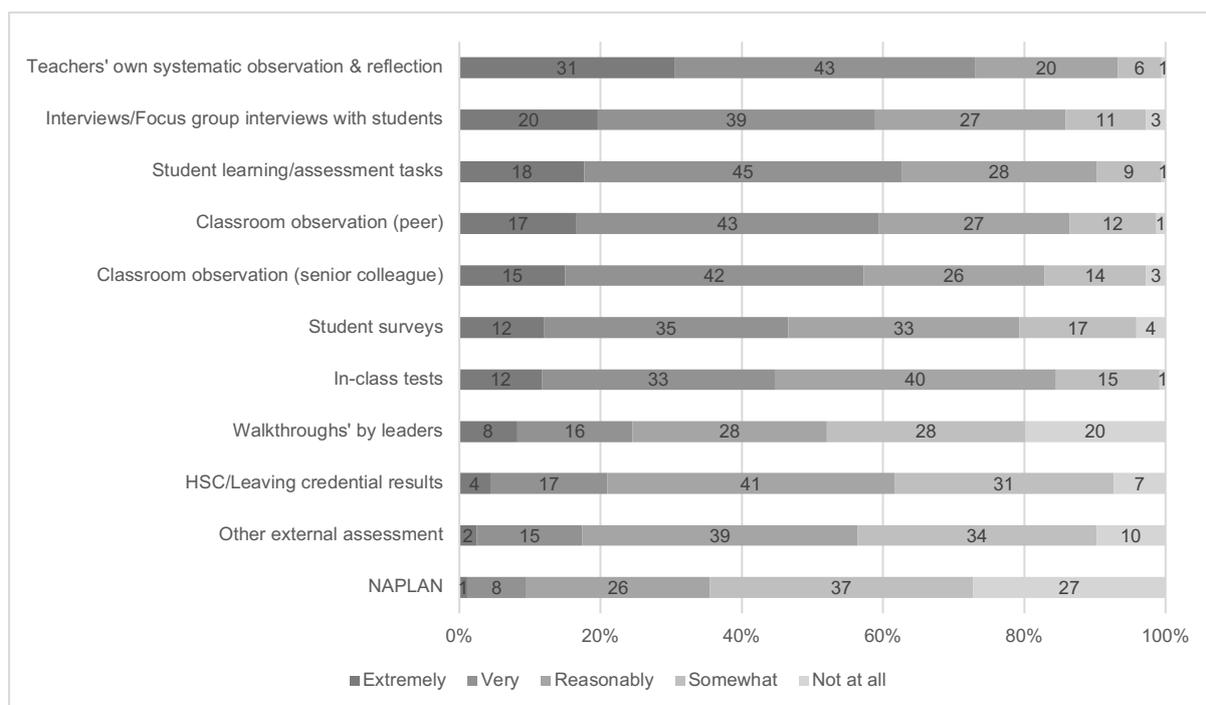


Figure 3: Valid and reliable sources of evidence of practice

The evidence most highly valued by this sample of enthusiasts was that generated via teachers' own systematic observation and reflection, where 94% indicated that they regarded this source of evidence to be extremely or very valid and reliable. This stands in contrast to the emphasis on randomised-controlled trials that has emerged in various contexts in recent years (see e.g. Goldacre, 2013). Student learning and assessment tasks, interviews/focus group interviews with students, classroom observation by a peer and senior colleague, student surveys and in-class tests were also seen as valuable sources of evidence, with at least 80% of participants indicating that they regarded these sources as either extremely, very or reasonably valid and reliable. Less trusted, however, were 'walkthroughs' by senior school staff, Higher School Certificate or leaving credential results, other external assessment data and NAPLAN data.

These perspectives were consistent with teachers' responses to the interview and open-ended survey questions about what constitutes good evidence of their teaching practice, where the most common responses related to indicators of student engagement. Figure 4 below highlights those codes allocated to at least 10% of responses garnered from participants (noting that responses were allocated as many codes as they reflected, and that some interrelate, e.g. engagement was sometimes described by participants as something that could be measured by student or collegial feedback).

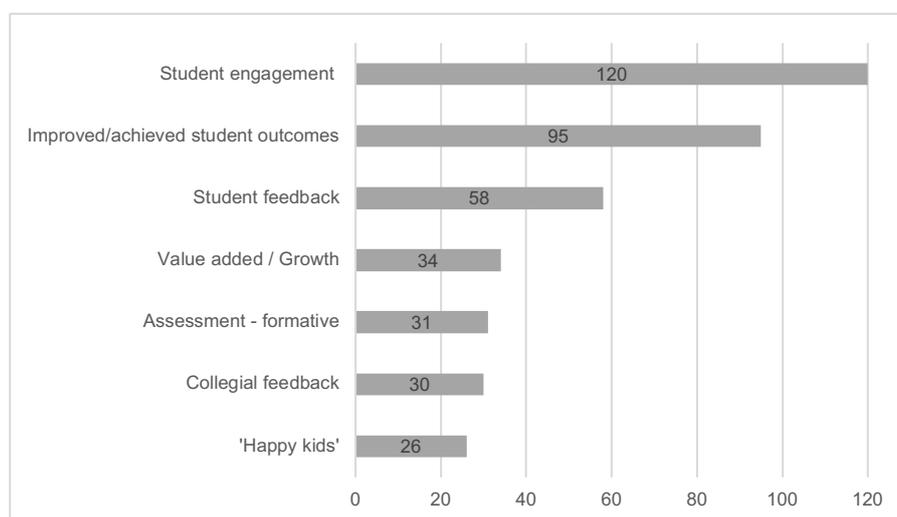


Figure 4: Evidence of practice (open-ended responses)

While student engagement in learning was the evidence most widely cited, student results, represented in achievement of or improvement against established learning outcomes and/or 'growth', including as measured by 'value added' were also cited by approximately 50% of the respondents. Feedback, both formal and informal from students, and formal feedback from colleagues was relatively highly cited, and over 10% of respondents nominated student formative assessment as a good source of evidence. Finally, approximately 10% of respondents gave responses linked to the in vivo code 'happy kids', a shorthand for happy students who want to come to class and learn, reminiscent of Munns and colleagues' concept of 'Big E engagement' (Munns & Sawyer, 2013, p. 19).

Importantly, very few of these responses related to a singular source of evidence; the vast majority (almost 90%) were complex and involved a number of interrelated sources. For example:

A variety of elements to ensure you get a holistic picture. Engagement of students, their feedback and their understanding. Their questions and results in a variety of activities. (Secondary Mathematics, NSW)

In another example, a respondent explained how they looked at:

...multiple layers of data including: student results, student improvement, feedback from other observing teachers, student surveys on engagement & feedback, HSC RAP² results, student work samples & my own personal reflection/analysis. (Secondary English, NSW)

In a more minor theme (about 4% of open-ended responses), good evidence was described as being ephemeral, and hard to pin down, with respondents using expressions such as 'the vibe' or 'flow':

When I am teaching well I have a physical sensation of 'weaving' ideas and concepts into a framework and there is flow. (Secondary English, NSW)

² The Higher School Certificate Results Analysis Package (RAP) is provided by the NSW Education Standards Authority to all schools to enable analysis of students' HSC results each year.

While the number expressing this sentiment was relatively low in questionnaire responses, it was echoed in the more expansive responses that could be provided by interviewees. One participant, An, working as a Deputy Principal in a NSW Government school noted that for him, evidence of good teaching could be ‘fleeting, ethereal flash-in-the-pan and gone stuff’ which as he says, ‘you count that but you can’t data it’. In a further example, Saxon, a teacher in a NSW secondary school, articulated it in the following way:

I think that there would be a couple of things ... the kids, in my opinion, need to enjoy being in the class...the teacher needs to be able to make the content meaningful for the students' lives and make real-world connections and things like that.

I think that student achievement is obviously important as well. There are inevitably going to be outliers... but I think that achievement, if you are doing a good job at teaching the kids, then the results should reflect that, and I think that that's a quantitative element of the data.

I think as well, perhaps peer observations where another colleague can come in and observe and say whether you're meeting the standards and things like that, and for them to be able to provide evidence of that...I think that that would be another piece of data.

Here Saxon points to the importance of using a range of sources of evidence, gleaned from teachers’ own reflection on classroom interaction, from results that reflect student achievement, and from systematic observation by a colleague, to provide different ‘windows’ onto practice. His perspective, like that of Diana, a secondary teacher in a NSW school, who referred to this as ‘a triangulation of things’, reflects Mockler and Groundwater-Smith’s (2018) argument that true ‘evidence-based practice’ involves building complex portraits of practice to inform teacher decision making. While in the response of many other interviewees these different sources sat comfortably together, for others a more hierarchical relationship was in play, at times working as a source of conflict. Dorothy, a primary teacher in a Victorian Government school, for example, expressed a high level of scepticism about NAPLAN results and their current use:

At the moment, it's all about NAPLAN data, which is driving me absolutely bonkers. So basically now – and I know this is completely wrong - the teachers are now getting, like reviewed based on the NAPLAN data...breaking it down into what questions they got right and they're basically making the assumption that that teacher's not teaching it...

Like you end up spending the first two terms practising NAPLAN and then you get told ‘don't teach the test’ and then you get the test shoved in front of you and they're like ‘how come you're not teaching that, you need to teach that’, so there's this big contradiction...

I think too much emphasis on one [test] and planning and implementing certain programs due to one test isn't probably the best way for a school to operate.

At the same time, and somewhat contradictorily, Dorothy eschewed the idea that qualitative forms of evidence were reasonable or in fact that they constituted sound evidence at all:

You have to look at the numbers. You have to have that evidence because in the end you're going to have your [appraisal] and they're like ‘well why are you doing that’. You have to show something. You can't just go ‘well I think this’ because, sadly, it's not good enough to say ‘I think this’. You've got to have the numbers, you've got to have the proof, it's like with everything.

While Dorothy had talked at some length about her use of evidence to shape classroom planning and practice, the ‘real’ evidence that counted for her was that which could be used to demonstrate her performance to others. She saw the alternative to ‘the numbers’ to be ‘I think this’, pointing to her lack of confidence in and regard for teacher professional judgement, regardless of the sources of (non-

numerical) evidence on which it might be based. For her, ‘feelpinion’ was all that was left once ‘the numbers’ were removed, reinforcing critiques that current constructions of ‘evidence’ in education can be disempowering for teachers (e.g. Godfrey, 2017), with the ‘opinions of educators’ seen as irrelevant (Biesta, 2007, p. 2). In other words, while Dorothy recognised that evidence used adversarially was not helpful to teachers in terms of informing practice, she also felt deeply the necessity of gathering and using such evidence to performative ends.

Dorothy was not alone amongst our interviewees in this conflicted position, again reminiscent of Hardy and Lewis’ (2017) argument around the ‘doublethink’ of data and the associated positioning of teachers. Indeed, the tension between what participants believed to constitute good evidence of their practice and what they believed constituted acceptable evidence of practice in the eyes of others permeated the interview data. Qualitative data obtained via the questionnaire were necessarily less expansive and discursive than those obtained via interview. While many (68%) survey respondents nominated more than one source of evidence in their answer, these tensions were less evident, although responses did clearly reveal leanings toward more or less performative and/or intelligent forms of accountability.

Evidence and Accountability

In addition to the content analysis discussed above, survey responses to the open-ended question relating to evidence were additionally coded holistically on the basis of their alignment with performative and/or intelligent forms of accountability. This additional layer of coding complemented and also transcended the initial content analysis – here we were looking not only at the types of evidence nominated but the uses to which teachers seemed to be suggesting they might be put: whether the implied relationship between evidence and practice was a more adversarial or forensic one, thus aligning more with performative or intelligent accountabilities.

Measurable Improvement: Evidence for Performative Accountability

Of the 265 responses, a relatively small proportion nominated the kind of evidence that might be considered acceptable within a framework of performative accountability. 18 (7%) provided responses aligned entirely with performative accountability, while a further 57 (22%) reflected both performative and intelligent forms of accountability in their nomination of evidence. These responses tended to rest on evidence generated through student results, improvement over time as reflected in pre- and post-test results, value-added growth and so on. The language of measurement and improvement permeated these responses, and the sources of evidence nominated were typically quite narrow in scope.

For some of these responses, such evidence could only be generated externally, such as in the case of credentialing examinations:

Can only tell after repeatedly teaching VCE [Victorian Certificate of Education] classes with external assessment calibrated against the GAT [General Achievement Test]. (Secondary Science, Victoria),

while for others, such evidence could be generated at school level using an appropriate method:

Pre and post testing informs whether you have made an impact in learning. The data comparison will inform how you modify practice to achieve learning gain. (Secondary English, NSW)

For other responses concerned with measurement of results, however, such evidence was about demonstrated improvement or growth, measured either in relation to established benchmarks or against the performance of others:

Students reach benchmarks, improved percentage of students at C or above, improved progress. (Secondary Health & Physical Education, Queensland)

Student growth outpaces the state/other classes in the school. (Secondary Mathematics/ Science, Victoria)

Here we see resonance with Holloway and Brass' (2018, p. 379) 'marketized, managed and performative teacher', with an almost sole focus on numeric and countable data reflecting current contexts of datafication (as illuminated by, e.g. Power, 2004; Rose, 1991).

Many of these responses also relied heavily on the notion of 'data' as evidence, with 44% of responses coded entirely against performative accountability and 37% of responses coded against both performative and intelligent accountability explicitly naming data as a source of evidence (as opposed to only 5% of responses coded entirely against intelligent accountability). While 'student data' was often invoked, others were more specific about the type of data that constituted evidence of practice:

Attendance and absconding data, assessment data, suspensions, N-awards. (Secondary Learning Support, NSW)

Improved data – pre and post test improvement. (Primary, Queensland)

All in all, the types of data invoked in responses that tended toward representations of evidence for performative accountability were highly quantified and represented measurable indicators of success said to be related to practice.

Learning, Engagement, Growth: Evidence for Intelligent Accountability

Of the 265 responses, 189 were coded as reflective of evidence for intelligent accountability, while, as noted above, 57 included such representations alongside those for performative accountability. These responses typically represented sources of evidence of practice to be complex and multiple, often explicitly reflective of good relationships within the classroom and reliant on teacher professional judgement and interpretation. For example,

When lessons are designed in response to my knowledge of how each of my students is understanding, and performing in an area, and are tailored to build/extend/improve/deepen their skills and understanding. When teaching these lessons, my students are visibly interested, showing signs of interest, some excitement for taking risks or developing their understanding, and confidently attempting the tasks provided. When the feedback from my students is that they are feeling positive about that learning experience, that they are more confident in that area, and increasing their understanding, and when the work they produce in that lesson demonstrates that they are developing the skills and understandings that I have intended for them through my planning. A really important aspect of this is that it occurs in a relational context. (Primary, NSW)

I know I'm teaching well based on how well my students synthesise their knowledge and readily apply it in different contexts. Also by the quality of their questions they ask me and each other in class. They come prepared to debate. Also when they help each other and are not afraid to take risks. When they send me essays and ideas they might be thinking about. Essentially I know I'm teaching well because the relationship is positive and students can articulate what they're doing, why they're doing it and can also show they understand, by teaching their peers. (Secondary English, NSW)

For many of these teachers, evidence of good practice is seen to be highly contextual and contingent, and the capacity to make judgements about what such evidence shows relies on good knowledge of

their students and positive learning relationships. This counters narrower forms of ‘evidence-based’ practice in which context is largely denied (Biesta, 2007). Designing lessons to build on prior understanding or making judgements about the quality of questions asked relies to a great extent on teachers’ knowledge of their students as learners: their current and background knowledge, their interests and capacities, and such knowledge rarely develops independent of good classroom relationships (see e.g. Comber, 2006). Indeed, more than half of the 120 responses focused on student engagement pointed to the importance of context and relationship in teachers shaping the conditions for such engagement to occur.

When the students are actively engaged because you have scaffolded the learning to build on current known knowledge in small do-able amounts that challenge them but are also relevant to their world. They understand when explicitly taught and can go away and independently practice what has been taught in fun and engaging ways to build strength [and] understanding on what has been learnt. Can explain their reasoning and observe them doing the required tasks well. (Primary, NSW)

Typically, these responses referred to a variety of different types of evidence. For such teachers, sources of evidence were used together in an ongoing, reflexive way, to shape and develop practice, in context:

I collect data in a variety of ways to try and figure that out. I collect the results of students, I do self-reflection surveys and I use students’ evaluations. All that is analysed. I also seek verbal feedback and input from students during the term to see if I am teaching in a way that suits their needs and engages them. I try my best to adjust what I am doing in the classroom if it needs changing and then get additional feedback. (Secondary Visual Arts, NSW)

In these responses, teachers point to a position on data and evidence akin to Cochran-Smith and Lytle’s (2009) notion of ‘inquiry as stance’, ‘a larger epistemological stance, or way of knowing about teaching, learning and schooling’ that ‘talks back to, or challenges’ the use of narrow sources of evidence (p. 44).

The responses relating to evidence for intelligent accountability highlighted forms of evidence reliant on rather than eschewing teacher professional judgement for interpretation and connection through to practice. Triangulation of different forms of evidence reflective of high-quality learning relationships and derived both as part of ‘business as usual’ (such as formative assessment) and more systematic data collection (such as ‘exit slips’) was commonly reported.

Conclusion

The privileging of particular kinds of educational research and data as evidence of teaching practice can feed into teachers’ understandings of themselves and their work. In this research, we found evidence of a similar kind of ‘doublethink’ to that identified by Hardy and Lewis (2017), with ‘data’ identified as something that must be ‘cared’ about more than ‘research’, suggesting its politicised status as a perceived, performative imposition on teachers, even within this generally enthusiastic sample of ‘teacher nerds’ and ‘data queens’. This suggests some ambivalence regarding teachers’ understanding of their work and how it is valued, or not, by the system. As such, the results of this study indicate that there may be an important and productive gap in play for Australian teachers between current forms of knowledge about education privileged in policy, and teachers’ own professional knowledges.

As noted earlier, however, the growing importance placed on narrow, quantifiable conceptualisations of evidence, and the related rise in performative accountability regimes, is not limited to Australia. Malone and Hogan (2020) argue that a recasting of ‘teacher professionalism’ on an international scale is taking place, led by three trends: the increased control over teachers’ work exercised by

transnational organisations such as the OECD; the increased purchase of ‘data’ in conceptualisations of the ‘good’ or ‘professional’ teacher; and the ongoing use of de-contextualised value-added measures in evaluating teachers’ work. Each of these three trends is both a consequence and a means of expanding the reach of narrow conceptualisations of evidence, and the teachers in our study demonstrated some of the complex ways in which the second of these trends in particular is playing out in Australia.

In terms of teacher response, Holloway et al. (2017, p. 8), in summarising contributions to a special issue on teacher accountability and high-stakes assessment, argue that, internationally:

either teachers do not approve of the policies (New Jersey and Rio de Janeiro), or they understand them differently than intended by the reformers (Shanghai), or they are left out of the policy formation process (Mexico), or they seek to ‘talk back’ to policy-makers by engaging with alternative and ‘richer’ forms of accountability (PISA for Schools in the US).

Along similar lines in her work on teacher resistance, Santoro (2016) has highlighted how narrow conceptualisations of evidence proliferating in the US in the form of ‘what works’ have underpinned moves to ‘teacher proof’ the classroom through the trojan horse of ‘fidelity of instruction’. She offers a rejection of such fidelity – often in the form of the rejection of these narrow forms of evidence – as an example of teachers providing an ‘intelligent response’ (Dewey, 1916) to regimes of performative accountability. These observations resonate with our findings in Australia, where, we argue, a similar sense of ‘talk back’ or ‘intelligent response’ is evident. A contribution of this article is that it places discussion of accountability overtly within the context of teachers’ own conceptualisations of ‘evidence’ in teaching, rather than primarily addressing high-stakes forms of teacher assessment. Opening up discussion of evidence – along with the role of teacher professional judgement – is particularly important in climates where evidence has come to be understood in narrow, limited and limiting ways, as we argue it has been in Australia and beyond (including in England, see Furlong et al., 2014; Godfrey, 2017).

In the first half of this article, we argued that there was significant warrant for this research given its potential to reveal productive ways forward in navigating teacher evaluation. The results of this Australian study contribute to wider international debates to indicate that we may be at a critical juncture in this endeavour. The data we have presented show that teachers’ discussion of evidence demonstrates the existence, and possible futures for a more ‘intelligent accountability’ in teaching, with teachers articulating the possibilities of ‘intelligent accountability’ in ways that current policy approaches rarely emphasise. Teachers have rich and complex approaches to their work and how they seek to make sense of it, relying on a range of evidence and, although subject to particular accountability pressures around what can be ‘counted’, also perceiving that this was not and should not be all that ‘counts’. Policymakers both in Australia and elsewhere would do well to heed the voices of the teachers presented here, and their ‘intelligent’ conceptualisations of what constitutes ‘good’ evidence of teaching work.

Biographical Notes

Nicole Mockler is an Associate Professor in the Sydney School of Education and Social Work at the University of Sydney. Her research focuses on education policy and politics, particularly as they relate to teachers’ work.

Meghan Stacey is a lecturer in the sociology of education and education policy in the School of Education at the University of New South Wales. A former high school teacher, she takes a particular interest in education policy, teachers, and the operation of dis/advantage within systems of schooling.

References

- Australian Bureau of Statistics. (2020). *School, Australia, 2019*. Retrieved from <https://www.abs.gov.au/ausstats/abs@.nsf/mf/4221.0>
- Ball, S. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215-228. <https://doi.org/10.1080/0268093022000043065>
- 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. <https://doi.org/10.1111/j.1741-5446.2006.00241.x>
- Braun, A., & Maguire, M. (2018). Doing without believing - enacting policy in the English primary school. *Critical Studies in Education* 61(4), 433-447. <https://doi.org/10.1080/17508487.2018.1500384>
- Cain, T., & Allan, D. (2017). The invisible impact of educational research. *Oxford Review of Education*, 43(6), 718-732. <https://doi.org/10.1080/03054985.2017.1316252>
- Cochran-Smith, M., & Lytle, S. (2009). Teacher research as stance. In B. Somekh & S. Noffke (Eds.), *The SAGE handbook of educational action research* (pp. 39-49). Sage.
- Comber, B. (2006). Pedagogy as work: Educating the next generation of literacy teachers. *Pedagogies: An International Journal*, 1(1), 59-67. https://doi.org/10.1207/s15544818ped0101_9
- Dewey, J. (1916). *Democracy and education*. MacMillan.
- Furlong, J., Menter, I., Munn, P., Whitty, G., Hallgarten, J., & Johnson, N. (2014). *Research and the Teaching Profession: Building the capacity for a self-improving education system*. BERA.
- Godfrey, D. (2017). What is the proposed role of research evidence in England's 'self-improving' school system? *Oxford Review of Education*, 43(4), 433-446. <https://doi.org/10.1080/03054985.2017.1329718>
- Goldacre, B. (2013). *Building Evidence into Education*. Retrieved from London: <https://www.gov.uk/government/news/building-evidence-into-education>
- Groundwater-Smith, S., & Mockler, N. (2009). *Teacher Professional Learning in an Age of Compliance: Mind the Gap*. Springer.
- Hardy, I., & Lewis, S. (2017). The 'doublethink' of data: Educational performativity and the field of schooling practices. *British Journal of Sociology of Education*, 38(5), 671-685. <https://doi.org/10.1080/01425692.2016.1150155>
- Hardy, I., Reyes, V., & Hamid, M. O. (2019). Performative practices and 'authentic accountabilities': Targeting students, targeting learning? *The International Education Journal: Comparative Perspectives*, 18(1), 20-33.
- Hattie, J. (2012). *Visible learning for teachers: maximizing impact on learning*. Routledge.
- Helgetun, J., & Menter, I. (2020). From an age of measurement to an evidence era? Policy-making in teacher education in England. *Journal of Education Policy*, 1-18. Advance online publication. <https://doi.org/10.1080/02680939.2020.1748722>
- Holloway, J., & Brass, J. (2018). Making accountable teachers: The terrors and pleasures of performativity. *Journal of Education Policy*, 33(3), 361-382. <https://doi.org/10.1080/02680939.2017.1372636>
- Holloway, J., Sørensen, T. B., & Verger, A. (2017). Global perspectives on high-stakes teacher accountability policies: An introduction. *Education Policy Analysis Archives*, 25(85). <http://dx.doi.org/10.14507/epaa.25.3325>
- Ladwig, J. (2018). On the Limits to Evidence-based Learning of Educational Science. In G. E. Hall, D. M. Gollnick, & L. F. Quinn (Eds.), *The Wiley Handbook of Teaching and Learning* (pp. 639-659). John Wiley & Sons.
- Leat, D., Reid, A., & Lofthouse, R. (2015). Teachers' experiences of engagement with and in educational research: what can be learned from teachers' views? *Oxford Review of Education*, 41(2), 270-286. <https://doi.org/10.1080/03054985.2015.1021193>
- Lingard, B., Sellar, S., & Lewis, S. (2017). Accountabilities in schools and school systems. *Oxford Research Encyclopedia of Education*. Retrieved 16 Sep. 2020, from <https://doi.org/10.1093/acrefore/9780190264093.013.74>

- Lyotard, J. F. (1984). *The Postmodern Condition: A Report on Knowledge*. University of Minnesota Press.
- Lysenko, L., Dagenais, C., Abrami, P., Bernard, R., Janosz, M., & Pigeon, M.-E. (2007). Questionnaire on the Utilization of Research-based Information. Concordia University. https://www.concordia.ca/content/dam/artsci/research/csip/docs/QURBI_Van_29_01_07_full.pdf
- Malone, A., & Hogan, P. (2020). Evidence and its consequences in educational research. *British Educational Research Journal*, 46(2), 265-280. <https://doi.org/10.1002/berj.3580>
- McKenzie, P., Weldon, P., Rowley, G., Murphy, M., & McMillan, J. (2014). *Staff in Australia's schools 2013: Main report on the survey*. ACER.
- McKnight, L., & Morgan, A. (2020). A broken paradigm? What education needs to learn from evidence-based medicine. *Journal of Education Policy* 35(5), 648-664. <https://doi.org/10.1080/02680939.2019.1578902>
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation* (2nd ed.). Jossey-Bass.
- Miles, M. B., Huberman, A. M., & Saldana, J. (2013). *Qualitative data analysis: A Methods Sourcebook* (3rd ed.). Sage.
- Mockler, N., & Groundwater-Smith, S. (2018). *Questioning the Language of Improvement and Reform in Education: Reclaiming Meaning*. Routledge.
- Munns, G., & Sawyer, W. (2013). Student engagement: The research methodology and the theory. In G. Munns, W. Sawyer, & B. Cole (Eds.), *Exemplary Teachers of Students in Poverty* (pp. 14-32). Routledge.
- O'Neill, O. (2002). *A question of trust: The 2002 Reith Lectures*. Cambridge University Press.
- O'Neill, O. (2013). Intelligent accountability in education. *Oxford Review of Education*, 39(1), 4-16. <https://doi.org/10.1080/03054985.2013.764761>
- Power, M. (1994). *The Audit Explosion*. Demos.
- Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford University Press.
- Power, M. (2004). Counting, control and calculation: Reflections on measuring and management. *Human relations*, 57(6), 765-783. <https://doi.org/10.1177/0018726704044955>
- Power, M. (2007). *Organized Uncertainty: Designing a World of Risk Management*. Oxford University Press.
- Prendergast, S., & Rickinson, M. (2019). Understanding school engagement in and with research. *The Australian Educational Researcher*, 46(1), 17-39. <https://doi.org/10.1007/s13384-018-0292-9>
- Rose, N. (1991). Governing by numbers: Figuring out democracy. *Accounting, organizations and society*, 16(7), 673-692. [https://doi.org/10.1016/0361-3682\(91\)90019-B](https://doi.org/10.1016/0361-3682(91)90019-B)
- Santoro, D. A. (2016). "We're not going to do that because it's not right": Using pedagogical responsibility to reframe the doublespeak of fidelity. *Educational Theory*, 66(1-2), 263-277. <https://doi.org/10.1111/edth.12167>
- Talbot, D. (2016). Evidence for no-one: Standards, accreditation, and transformed teaching work. *Teaching and Teacher Education* 58, 80-89. <https://doi.org/10.1016/j.tate.2016.05.006>
- Torrance, H. (2014). Philosophy Is Not Enough: Inserting Methodology and Politics into the Space Between Science and Democracy. In A. Reid, E. P. Hart, & M. Peters (Eds.), *A Companion to Research in Education* (pp. 401-405). Springer.
- Verger, A., Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: A political sociology of global education reforms. *Educational Review*, 71(1), 5-30. <https://doi.org/10.1080/00131911.2019.1522045>
- Welch, A. (2015). Evidence-based policy: Epistemologically specious, ideologically unsound. In H. Proctor, P. Brownlee, & P. Freebody (Eds.), *Controversies in education: Orthodoxy and heresy in policy and practice* (3 ed., pp. 63-80). Springer.
- White, S., Nuttall, J., Down, B., Shore, S., Woods, A., Mills, M., & Bussey, K. (2018). *Strengthening a research-rich teaching profession for Australia*. Australian Teacher Education Association (ATEA), Canberra; Australian Association for Research in Education (AARE), Melbourne; Australian Council of Deans of Education, (ACDE) Canberra.

