



Dr. Ian Hardy shares his research into some familiar school-based professional development practices...

Introduction

This article explores how continuing pressure for improvements in standardised tests, and associated measures of student learning in the Australian state of Queensland, influenced teacher learning practices. The effects of high-stakes testing have been expressed more overtly vis-à-vis schooling practices in some national settings, most obviously the United States, where national and state-based standardised tests have been deployed to make decisions about school closure/reconstitution, staff termination and student grade/year level retention (Mehta, 2015).

However, the specific effects of such tests in relation to teachers' learning practices in other settings, under current conditions, is an area for further inquiry. This includes specific formal, ongoing instances of teacher professional development practices, such as those described here. Furthermore, exploring whether and how such instances of teacher learning can be sites for the cultivation of perverse effects in schooling settings, including in relation to teachers' data literacy, is an area for further investigation. The research presented in this paper seeks to reveal the nature of the tensions and points of possibility for the teacher learning that transpired in one school in northern Queensland, Australia, in the context of increased attention to standardised testing and associated data.

To provide insights into the effects of standardised and quantified measures of student learning on teachers' learning, the research draws upon the discussion that characterised a collaborative inquiry process – described as 'Inquiry Cycles' – involving teachers meeting together in individual year level groups for a full day per term to discuss evidence of students' learning as a vehicle for their own learning for improved practice. The research focuses upon a group of Year 3 teachers, and draws upon an eclectic 'toolbox approach' (Ball, 1994) to educational policy sociology to make sense of these governing practices, and teachers' responses to such practices. This includes multiple theoretical resources, including processes of governing through numbers, and applications to the field of education.

See Attachment 1 for Abstract, Conceptual Resources and Literature Review



Context

National testing in Australia

In the context of a global neoliberal policy imaginary in education (Rizvi & Lingard, 2010), the focus upon global indicators of literacy/reading capacity has had a variety of impacts at the national level. In the Australian case, the former Education Minister (and subsequently, Prime Minister), Julia Gillard, and Prime Minister Kevin Rudd justified the introduction of standardised national literacy and numeracy testing in 2008 as part of a strategy to improve Australia's educational performance as a means of enhancing its broader economic performance (Gable & Lingard, 2013). This was augmented by Prime Minister Gillard's subsequent call to catapult Australia into the top 5 nations on PISA results by 2025 (Grattan, 2012). However, while the broader global infrastructure of numbers referred to in Attachment 1 exert influence at the national level – as evident in the various 'PISA-shocks' attending specific countries' results on these 'global' measures (e.g. Germany as a result of its initial participation in PISA in 2000, and Japan in 2003 (reading literacy) and 2006 (math literacy) (Wiseman, 2013)) – it is the range of commensurate measures at the national and sub-national levels which have had more immediate effects upon student and teachers' learning practices.

Since 2008, standardised national literacy and numeracy tests – NAPLAN – involving students in Years 3, 5, 7 and 9, have been undertaken in Australia. In the inaugural tests, students in Queensland achieved poorly relative to students in most other states. A strong, critical media focus on Queensland students' outcomes led to renewed policy attention to educational practices in schools. At the same time as the testing regime was implemented, a simultaneous focus upon developing a national curriculum was under way. Amidst ongoing concerns about students' outcomes on these tests, Queensland became one of the first states to trial the new Australian Curriculum, developing its own detailed version of the Australian Curriculum – 'the Curriculum into the Classroom' ('C2C').

Teachers' learning at 'Lillyvale'

As a result of concerns about Queensland's performance on NAPLAN, many schools and educational districts/regions also required teachers to collect data in the form of additional standardised tests (particularly 'Progressive Assessment Tests¹ – Reading (PAT-R)', and 'Progressive Assessment Tests – Mathematics (PAT-M)', as well as reading level data in the form of students' attainment on the 'PM Benchmark'² levelled reader series.

¹ The Progressive Assessment Tests are standardised tests designed to measure improvement in students' results in generic literacy (reading; vocabulary) and numeracy capabilities, typically over a one year period. Students sit the test at the beginning of the year, and again at the end, with the difference between the results indicating improvement (or regression) over that time.

² PM Benchmarks are a levelled reading series designed to ascertain students' capacity to read and comprehend texts. These levelled readers are typically used in the early years (Preparatory to Year 3) in the Australian context.



Within this broader context of concerns about results, particularly in the northern regions of Queensland, a decision was made at 'Lillyvale' Primary (Preparatory to Year 6), a school located in one of these northern regions, to undertake a more systematic, ongoing professional learning approach to teachers' learning to try to encourage more substantive teacher learning for student learning. As part of this process, the Head of Curriculum at the school, who had previously been a regional literacy consultant, advocated for the development of a series of ongoing meetings, or 'Inquiry Cycles', on the part of teachers to enable them to more systematically inquire into their practice. Influenced by the work of literacy educator Helen Timperley (2009), this inquiry approach was supported by the principal, and after some difficulties establishing the meetings, from January 2013, teachers met together as a year level group one day per term to inquire into the nature of their teaching practices. This involved analysing units of work teachers were about to teach, as well as interrogating evidence of their students' learning from completed units to inform future teaching practice. This evidence included curriculum data (on an 'A to E' scale) associated with the C2C, but also other data in the form of national testing results (NAPLAN data), various 'reading level' data (PM Benchmark results) and other school-collected standardised literacy and numeracy data (PAT-R; PAT-M).

See Attachment 2 for Methods and Methodology

Findings: Governing teachers' learning

As a result of this process, three broad themes were identified within the data. These pertained to the governing of teachers' learning through:

- teachers grappling with the focus upon achieving particular reading levels by the end of each year;
- the use of standardised, numeric data, including reading levels, in hierarchical relations with NAPLAN at the pinnacle, and the use of these numbers as a means of validating student results;
- and the use of students' results on standardised PAT-R reading data to measure the 'effect size' of teachers' pedagogies.

Each of these three themes is presented below, with further analyses provided in the subsequent discussion section.

Grappling with 'the number/s': The focus on reading levels

Even as teachers were ostensibly engaged in analysing students' samples of work, the governing of teachers' learning through numbers was evident in the way the middle part of the day began with teachers wanting to discuss the nature of their students' reading levels, including concerns about the value and validity of such levels/numbers. Based on the 'PM Benchmark' levelled reading series,



teachers were required to undertake a 'running record'³ every term to ascertain students' reading capacity/'levels' against the 30 identified levels in the reading series (from 1 (lowest) to 30 (highest)), and to record this number as evidence of students' reading capacity. This process constituted teachers' work as a source of frustration, with teachers questioning the extent to which the collection of such data was useful:

Lucy: Why do we apply a band aid effect to all children? Not all children are going to be able to read a text to you that's at a level that somebody came up with. If you hear them read all different texts all day long, which you should do in your classroom, you have a sense of how that child reads, what strategies they use, and all the other things you're supposed to know from that reading. I just don't understand this whole fascination with, 'They must be a 20', or whatever. And then you get parents coming in crying, 'Oh my God; my child's not a "20"!'

However, and in keeping with calls to try to develop professional capacity and language of instruction to improve teaching (Glazer & Peurach, 2015), the facilitator did not simply agree with this more reductive analysis around the reading levels. The way in which more educative logics were in contest with more performative logics was evident in her critique about the difference between focusing upon the educative potential of undertaking the PM reading levels, and the specific 'number' generated through this process:

Nova: But we've been trying to break down that thinking around PMs for years and years and years. You guys – like no offence anyone – but the teachers do it to themselves! We keep saying the PM-

Giselle: But we have to have the number to put on 'G-drive'⁴, we have to have it.

Nova: But it's not about the number; that's not the useful part. It's the information that you know!

Giselle: Yeah but it is the information that they want at Admin. Admin.⁵ want a number!

Lucy: And she [principal] told us it was systemic.

Nova: No – we want you to know what your students are capable of!

³ A 'running record' is a checklist of a student's reading capacity using a text that approximates that students' reading ability. Teachers mark, on a selection of text, the accuracy and fluency with which students read the text, and the extent to which they comprehend it.

⁴ 'G-drive' is a reference to the school's electronic administration portal.

⁵ The administration team ('Admin.') comprised the principal, two deputy principals, and the Head of Curriculum.



However, and revealing the contestation between more critical and compliant approaches to the governing of teachers' work and learning through this data, there was also an overt valuing of recording the number in the school's data repository to monitor consistency of teacher judgement. This valuing of the numbers in and of themselves is an instance of the perverse effects that characterize not only the work of systemic personnel (Lingard & Sellar, 2013), but, as evident here, the specific teacher learning practices of teachers in schools. The semiotics and valuing of this display – this visualisation of data (Williamson, 2015) – was explicit:

Giselle: So why do we have to put it on G-drive?

Lucy: And why do we have to use a PM to do that?

Nova: ... Because as a whole school we're having some consistency around the types of data we collect and use.

Giselle: But that's not consistent.

Nova: I know, Giselle, but that's what we're working towards. Why put it on? So we've got a visual.

Even as there was a focus upon 'what students are capable of', the semiotics of the 'numbers' and focus on 'consistency' indicated a normalisation of conduct through these numbers (Miller & Rose, 1990), and their display (Williamson, 2015). Reflecting more enumerative logics of practice (Author, 2015), teachers' work and learning were to be calibrated against these numbers, and these numbers would reflect the extent to which teachers were successful or otherwise. The concern for 'consistency' was a reflection of the need to orchestrate the work and learning of teachers and students, through close scrutiny of student data as evidence of teachers' work. At the same time, there were also intimations of processes of constant comparison as a distinctive mode of operation (Ozga, 2009, p. 150; Grek, 2009), and which would be enabled by portraying the numbers and data in the form of tables and ledgers to bring into being a more 'accurate' account and visual surveillance of teachers' work and learning, with the use of these visual displays of student results as evidence of this work and learning.

Consistency and 'in-the-head-judgements': The relationship to NAPLAN

Discourses of standardization were also evident in the way teachers' learning was oriented towards monitoring consistency in reading level data, particularly in relation to NAPLAN. This was apparent in teachers' concerns about how regional and state attention to NAPLAN were seen as contributing to demands for students to achieve at particular 'levels', even as the connection between the two was also contested by some teachers:



Molly: I think what needs to go is the expectation by the end of the year of what these kids have to be; that's what I think needs to go... [students] need to be a level 8. I think that needs to go ...

Nova: Nobody ever says that. We have said this over and over and over. The region did this regressional data where they worked back from Year 3. The state did it pretty much as well. So they look at the kids that are achieving in NAPLAN in Year 3, [and they] were reading '8' in Prep, '16' in [Year] 1, and '20' in [Year] 2⁶. ... So what if they were reading '8' in Prep and they did well in NAPLAN? What's the link? What's that got to do with it, I say?

This exchange reveals that at the same time as teachers' focus upon students attaining particular levels of achievement resonates with broader discourses of 'constant comparison' vis-à-vis data (Ozga, 2009; Grek 2009), the HOC's critique of such discourses represents a challenge to the technical rationality associated with NAPLAN as a form of 'base-line' data (cf. Comber, 2012), and more performative conceptions of NAPLAN as a vehicle for ranking rather than learning (Polesel *et al.*, 2014; Klenowski & Wyatt-Smith, 2012). She was not simply prepared to concede that these numbers as inscription devices were somehow productive of an accurate conception of student and teachers' work and learning, even as such numbers may be construed as constituting 'a navigable space of commensurability, equivalence, and comparison' (Lingard, Martino, Rezai-Rashti & Sellar, 2016, p. 2).

Nevertheless, there was also explicit recognition that the high stakes data associated with NAPLAN would be the principal determinants of teachers' success (or otherwise) with their students, even as other forms of evidence were seen as providing better insights into the actual 'story' of these teachers' student learning. This was the case even as it was recognised that teachers in Year 3 had spent considerably less time with these students than these students' Prep, Year 1 and 2 teachers:

Nova: At the end of the day the other sets of the data are going to tell the story aren't they? And I personally would be a little bit nervous about some of the other sets of the data if I was a Grade 3 teacher... you've only had the kids for a short amount of time by the time you're going to get judged for NAPLAN. I don't think that's fair; someone else has had them for 3 years!

The governing practices of the numbers that constituted the national testing regime were clearly evident through such processes of 'judgement', even as they were constituted as problematic. These numbers, and the processes of 'judging' teachers on the basis of these numbers, reflect the influence of performative discourses upon teachers' learning (cf. O'Leary & Wood (2016)) on the effects of such numbers as part of lesson observations), and how the terrors of performativity (Ball, 2003) that influence teachers' work under conditions of continual surveillance more broadly actually play out in practice. The focus on various numbers associated with high stakes testing reveals the more perverse

⁶ A reference to PM Benchmark levels students were expected to attain at each of the respective year levels.



effects associated with such testing, and how these play out in school settings – in this case in relation to a specific instance of teachers' learning. While there were efforts to foster more critical approaches to the focus upon the numbers for their own sake, teachers struggled to proactively respond to such invitations to engage critically with such numbers as data; this indicates how current discourses around the need to 'improve the data' in Australian schooling cultivate angst on the part of teachers around such numbers, but without substantive increases in uses of data, typically associated with more 'evidence-based practice' approaches.

Similar 'governance by comparison' (Grek, 2009; Ozga, 2009) practices were evident in how curriculum-generated assessment data were compared with standardised data as a matter of course, and the former judged against the latter. In one way, the reflection upon various forms of data, including curriculum data derived from students' work samples, reflects how the Inquiry Cycle days enabled teachers to come together to build upon collaborative capacities/understandings of student work to further develop their teaching practices, thereby potentially challenging more problematic renderings of standardised testing and its associated reductive effects on curriculum (cf. Pease-Alvarez & Thompson, 2014). However, that more standardised data were seen as a corrective to expressed concerns about teacher judgement also reveals a privileging of such data, and how NAPLAN has been seen to be more about 'ranking' and 'policing' than learning (Polesel *et al.*, 2014); it was the 'consistency' with NAPLAN data that was particularly valued:

Nova: What we do know, to take a bit bigger picture than the PMs, what we do know is that teachers have been making judgements about kids' progress based on 'in-their-head' knowledge. Now kids have been getting 'Cs' and 'Bs' and 'As'⁷ – there were a lot more 'Bs' and 'As' 5 years ago, I can tell you, than there are now! ... So the 'in-the-head' judgement is not consistent. ... The data [in other schools] won't match the NAPLAN data and ours does; pretty close. There's always a few anomalies.

Again, this continued reinforcement of the validity of standardised data, including using NAPLAN as part of a critique of 'in-the-head-judgements', indicates the normalisation of a broader process of governing by numbers (Miller & Rose, 1990). These numbers were seen as valid and valuable for the more 'authoritative' accounts of students' learning that they were seen to provide.

Going deeper? PAT-R and 'effect sizes'

Part of the solution to the dilemma of consistency was construed as the need to focus upon yet another form of standardised data – the Progressive Assessment Test for Reading (PAT-R) – as another measure of commensurability of teachers and students. As the discussion progressed, the suggestion was made that rather than focusing upon the PM levels to ensure consistency of evidence of student

⁷ These letters refer to students' results (from A to E) in their regular subject areas (described as Key Learning Areas ('KLAs')).



learning, greater emphasis could be placed on PAT-R results, and the 'effect size' of students having been in particular teachers' classrooms (as measured by the change in students' results on the same PAT-R test from the beginning to the end of the year).

The power of this enumerative discourse was evident in how concerns about the accuracy of PAT-R results were seen as assuaged by referring simultaneously to other forms of data, including NAPLAN results:

Lucy: Do you think that the PAT-R data is a true indication of their ability, though, 'cause I've got some that are very 'iffy'⁸.

Coralie: This first pre-test – okay you've got to remember it's a year above them okay. Usually what we try to look for is if they're good readers, they should come out at about average or so, or above in their PAT-R. That should also be reflected in their KLA marks and their NAPLAN marks. They usually sort of align.

This exchange led into a discussion about the change in students' results on the PAT-R test from the beginning of the year to the end – the 'effect size'. The power of these numbers, reflective of the dominance of benchmark data more broadly (Datnow & Hubbard, 2015), was evident in the way in which they were valued by the HOC as a mode of measuring the progress (or otherwise) of students, even as they constituted a source of significant angst amongst the teachers:

Nova: And what we do know when you do it across a whole school it becomes really obvious when some things are working really well in some classrooms, and some are not. And, of course, you're going to get anomalies; you're going to have kids that are going to shift your effect size from year to year. If you've got eight kids that are three years below or something like that, you can explain those kids. But there have been some classes where none of the kids are progressing.

...

Jean: Seriously, it's terrifying; you seriously lose sleep over it.

Lucy: Poor Dorothy – she had a meltdown last year.

These 'effect sizes' appeared to take on a life of their own. They were construed, at times, as if they were the primary focus of attention, and something to which teachers needed to be able to respond to justify their practice. In this sense, the effect sizes became the target of teacher concerns, and those classes in which there was a failure to be able to show adequate improvement in effect size could be construed as 'classes where none of the kids are progressing'. Again, perverse effects were not simply expressed systemically (Lingard & Sellar, 2013), but at the level of the school, and as flagged in the

⁸ Uncertain.



research presented here, in relation to specific instances of teachers' learning. Teachers' strong focus upon effects sizes in and of themselves represent how attention to particular kinds of standardized data can actually reduce teachers' data literacy as they become obsessed with such data as abstracted indices of teacher quality. Disconnected from any understanding of what produces these results, such foci thereby encourage much more reactive and performative responses on the part of teachers – responses designed to increase the effect size 'number', rather than active engagement around students' learning. Reflecting the dominance of these markers of achievement, teachers could recite their specific effect sizes. Also, and as evidence of the constant comparison that characterises the governance of education through numbers (Grek, 2009; Ozga, 2009), and the influence of enumerative logics more broadly (Author, 2015), teachers were aware of the 'numbers' that they should be attaining:

Sandy: What was your effect size last year Lucy?

Lucy: Not much: .49.

Jean: 1.22

Kim: .51...

Sandy: .4 is bare minimum, .7 is our goal...

Such responses reflect the power of these numbers in this context, and how this focus on these numbers constituted this collaborative discussion; this dialogue is reflective of the effects of broader audit discourses that characterize so much policy-making associated with teacher learning in Australia at present (Mockler, 2013). Teachers' learning, through this collaborative discussion, was a site for the construction of the PAT data as a form of valuable standardised data, even as it was recognised as problematic; this occurred alongside the simultaneous constitution of NAPLAN as a 'consistent' marker of achievement:

Nova: See, it's hard with you guys because of the sets of data that we haven't got. So traditionally you've got nothing that's consistent until Grade 3 with the NAPLAN. So you had PMs you were doing, but we knew they were all over the place, and probably always have been. So then we started doing the PAT-R, and there's still issues with that; it's not the best set of data but it's something.

Molly: Well, yeah, especially when my little boy Toby said to me, in his PAT-R, 'I just copied all the answers off the girl next to me!' ... And I just thought how many kids are doing that⁹?

The PAT data were valued because they were seen as providing a form of standardised data considered more scant in some year levels, including Year 3. This was the case even as there was also evidence of an epistemic community (Glazer & Peurach, 2015) evolving amongst these teachers which was aware of, and reflected on, how the numbers were sometimes clearly inaccurate.

⁹ The classroom teacher was not present when her students sat the PAT-R test.



In spite of these concerns however, including not knowing what to attribute the PAT data improvements to, the governing of teacher learning through PAT data monitoring was clearly evident in the discussion and explicit support for what such data might indicate about particular teachers' pedagogical practices:

Nova: It's only one set of data; you don't know what to attribute the effect to.

Sandy: No, that's what I found out.

Molly: This one - what Toby said - that was amazing; I would never have thought of that.

Nova: Yeah. So there's a whole lot of things that are happening with the kids. But when you get a whole class that's doing really well, even a couple of years in a row, from an Admin. point of view, you'd be thinking, 'Hang on a minute; what's going on in that class? Why are the kids in that class achieving more than kids in another class? What's working well?'

Again, even as there was an explicit understanding of the problems associated with emphasising these numbers, these numbers were productive of particular (enumerable) conceptions of practice, and clearly valued for their capacity to monitor student learning, but particularly as tools for evaluating teacher performance. The explicit connection between 'what's working' and substantial improvements in PAT results 'even a couple of years in a row' reflects the influence of these numbers, and how the deployment of such numbers can serve to constitute particular kinds of communities of teacher learners, downplaying other forms of evidence, and broader conceptions of professional judgement more generally (Piro & Mullen, 2013).

Discussion: Understanding teachers' compliance with and critique of standardization

The data provide glimpses into the complexity of teachers' learning at this school site. Specifically, the Inquiry Cycle discussions are powerful for how they reveal the governance of teachers' learning under current policy conditions, manifest in the form of both teachers' compliance with and critique of a strong focus upon school, regional, state and national data. How this complexity is rendered 'knowable' is particularly important for understanding how the work and learning of these teachers was practised. This sense making was evident through teachers' responses to various forms of standardization of practice – expressed through a myriad of numbers – and brought into being through the teacher exchanges that constituted the Inquiry Cycle approach. At all times, teachers' discussions revealed how this work was understood through the plethora of numbers, a logic of enumeration (Author, 2015), that characterised attempts to 'capture' their students' learning, and by proxy, what they themselves had (or had not) learnt as teachers. Teachers came to see how their work was made knowable through the various enumerative techniques that characterised summations of their students'



learning, and, by association, their own learning. Importantly, this engagement with enumerative conceptions of student learning led to various perverse effects in relation to data use/literacy at the school level, including in relation to more formal, long-term instances of teachers' learning – an area not well represented in the literature.

The complex effects and power of these standardization practices were manifest in various instances of teachers' responses to these numbers. This was clearly the case in relation to the PM Benchmark data. Teachers' work and learning were constituted as an administrable domain through the way in which students' reading comprehension could be distilled to a single number through the PM Benchmark reading series. These teachers were learning, through the talk that constituted the Inquiry Cycles, that this number did not simply relay information about the nature of students' learning, but simultaneously about teachers' teaching as well. The focus upon 'the number' in relation to the levelled reading process reflects how these numbers work to 'shape, normalise and instrumentalise' (Miller & Rose, 1990, p. 8) the way in which teachers make sense of their work.

This was the case even as the arbitrariness of the reading levels was critiqued – numbers 'that somebody came up with'. The broader 'trust in numbers', with its associated 'objectivity' (Porter, 1995), was certainly challenged by teachers seeking to focus greater attention on the variety of texts to which students were exposed on an ongoing basis as part of their classroom practices; in this sense, power was not simply repressive but also productive of teachers seeking to focus upon the nature of their students' actual reading practices, rather than ways of 'measuring' reading capacity. However, even as teachers were critical of the demands to collect standardised measures of their students' reading attainment via PM Levels, and contested such an approach as inadequately focused upon the nature of students' actual reading practices and the rich array of resources they read as part of their everyday work, this critique was always necessarily undertaken in the context of – in relation to – these enumerative technologies of control and comparison (Grek, 2009; Ozga, 2009). That teachers were debating the nature of the actual PM Benchmarks as part of their discussion reveals that even as there was a lack of consensus about the specificity of this technique, teachers' learning, in the form of the Inquiry Cycle discussions, was clearly governed by the deployment of these numbers, and this focus on the numbers had perverse effects at the school level.

The PM Benchmarks constituted particular indices of learning as part of teachers' individual and collective development – indices overtly recognised as problematic if used for more performative rather than educative purposes. Teachers' conduct was disciplined through this emphasis upon the numbers, even as the numbers were contested. Teachers' references to 'Admin. want a number' indicated how they were clearly responsive to demands that they make their work and learning calculable to those in authority in the school, and that they calculate about such learning (Rose, 1991; Simola *et al.*, 2011); as in other instances of the enumeration of teachers' work and learning (e.g. numeric gradations of lesson observations (O'Leary & Wood, 2016)), more performative discourses were evident. And the reference to this being a 'systemic' request also reveals how these numbers were seen as pervasive within the schooling system in which these teachers worked, not simply their school. This concern with having to



produce these numbers reflects a broader culture of a 'tyranny of numbers' (Ball, 2015) in which such numbers may be understood as ascribed a value beyond their 'true worth', but also simultaneously reflective of a degree of 'worthiness' difficult to deny.

That a 'hierarchy' of numbers was in place was also evident in the way in which the reading levels were related to NAPLAN data, and the debate in which teachers engaged around consistency of student achievement, and how this might be best 'measured'. Again, even as the Head of Curriculum described how the regional and state efforts to fabricate associations between NAPLAN and reading levels were deeply problematic, and as she sought to appropriate the focus on testing for what she construed as potentially more educative purposes (cf. Author, 2014), such discussions reflect the influence of these enumerative foci. At the same time as the critique of the association between NAPLAN and reading levels was a challenge to processes of 'contextless accountability' that have been associated with such standardised assessment practices more broadly (Hutt, 2014), and even as participants worked hard to engage in broader processes of shared moderation of assessment practices against specified standards (Klenowski, 2011), the Inquiry Cycle was also a forum in which teachers came to learn that these numbers were seen as needing to be interrogated to better inform teachers' understanding of their practice. Such attention reveals how teachers' learning was clearly a site to be governed via state, regional and school-based data, problematic though this data may be.

The more overt compliance on the part of teachers to these numbers was evident in the way in which teachers recognised the school data were seen as more 'consistent' over time. This reflected a responsiveness to broader demands to ensure more careful deployment of the various modes of collecting evidence of student learning at the school level (PM Benchmark and PAT data in particular), and how these related to the broader systemic NAPLAN results. Such foci assume that more standardised data are relatively unproblematic in and of themselves – that NAPLAN is a form of 'base-line data' (Comber, 2012) amenable to improvement through, for example, coaching (see van Leent & Exley, 2013) – and that it is teacher consistency and judgment that are the primary concerns.

This prioritising of NAPLAN results is an example of the perverse effects of these more enumerative technologies in school settings, and specifically in relation to teachers' learning, and how teachers' data literacies can become adumbrated by more pervasive, performative discourses. Such responses do not give adequate regard to the problematic effects of NAPLAN more broadly (Carter, Klenowski & Chalmers, 2016; Klenowski, 2016; Klenowski & Wyatt-Smith, 2012), and that econometric modelling fails to account adequately for necessary professional judgment (Piro & Mullen, 2013). Also, the relatively poorer performance of Lillyvale Primary in A-E curriculum data when compared with other schools in the region, but consistency with NAPLAN results, reveals how a process of constant comparison is both the 'symbolic feature' and the 'distinctive mode of operation' of educational governance in schooling settings (Ozga, 2009, p. 150), and that more enumerative logics are central to this work (Author, 2015). The result is a tendency for more standardised measures to dominate over school-based judgements of student work more generally, and for teachers' data literacy practices to be skewed as a result.



This is in contrast with the sorts of epistemic communities of practice encouraged by Glazer and Peurach (2015) in which teachers' discussions could contribute to knowledge development, and as a corrective to more performative, enumerative practices. In a sense, the contestation around NAPLAN – evidenced, for example, through the HOC's critique of ascribing responsibility to those teachers whose students sit the test in Years 3, 5, 7 and 9 – could be seen as developing a particular epistemic community of teachers fostering more occupational control, and of efforts to try to respond more productively to various forms of the 'control society' (Thompson & Cook, 2014) represented by such testing. However, the extent to which such responses constitute modes of governance that challenge more reductive approaches to the use of standardised data that fail to adequately acknowledge valid and valuable site-based knowledge in classrooms and schools – the practical knowledge that is actually necessary for the successful functioning of classrooms (Labaree, 2011) – is a moot point in the context of the dominant technocratic rationality of NAPLAN in the Australian context (Comber, 2012). Pressures for improved levels on the benchmarks in relation to NAPLAN also reflect a form of 'adultification' of childhood, with developmentally inappropriate expectations of students imposed by adults ultimately feeling pressured to prove improvement on NAPLAN results (Bousfield & Ragusa, 2014).

Teachers' compliance with and critique of enumerative data as expressions of the governing of teachers' learning practices were also evident in teachers' focus upon improvements in effect sizes on the PAT-R tests. The focus upon the PAT-R results as another set of data to use in conjunction with PM benchmark and NAPLAN data to develop a 'deeper' understanding of students' capabilities, reflects a further expansion of the application of a technology of standardised numbers to try to 'capture' educational practice. Even when concerns were expressed about PAT-R results, there was a call to 'verify' this data further via other forms of standardized data; something of Berlant's (2011) 'cruel optimism' was evident in the way more learning-centred and profession-driven approaches, such as the Inquiry Cycles, were drawn upon to 'better' operationalize the focus on data. This governing by numbers was also evidenced through an elaborated discussion about the nature of the effect sizes in relation to student outcomes, and teachers' overall 'effect size'.

Teachers' ease of recitation of their specific effect sizes, sometimes to the second decimal place, and the way in which they compared these figures, reveal how these numbers were a significant technique deployed within the school – benchmark data as dominant (Datnow & Hubbard, 2015). The specification of a goal of .7, and minimum improvement of .4 seemed responsible for the 'terrors of performativity' (Ball, 2003) that literally attended this process ('it's terrifying; you seriously lose sleep over it'). Even though the focus upon PAT results, with their intrinsic value-added measures, were not operationalised at the national, state, regional or school levels such that they had the dramatic impacts associated with such value-added models in some parts of the United States, and in spite of the recognised gap between value-added accountability models of student achievement and actual student learning because of individual school, student and broader socio-political factors (Ready, 2013), the existence of such models actively governed how teachers' work and learning came to be constituted.



Again, more perverse effects were evident at not only the system level (Lingard & Sellar, 2013), but at the level of the school itself. Of particular concern was the way in which the effect size discussion appears to show that teachers' engagement with data actually seemed to be decreasing their data literacy, rather than enhancing it. The pervasive discourse around 'effect sizes' in Australian schooling played out in a way which meant these teachers became obsessed about their specific effect sizes, and how these numbers came to be understood as markers of teacher quality, rather than seeking to understand the conditions responsible for their production, and whether and how they may be valid sources of information about their students' learning. Such responses reflect how notions of professionalism have become increasingly 'datafied' (Thompson & Mockler, 2016), with increased potential for manipulation of such data as the stakes around such data increase (Thompson & Cook, 2014).

This is not to say that these particular techniques were necessarily always problematic, even as their limitations were clearly evident; this resonates with recent research indicating that aggregate results of NAPLAN can be useful for identifying areas for professional learning (Polesel *et al.*, 2014). Even as standardised tests and other forms of standardised learning only provide limited information in comparison with alternative forms of documentation in assessment (Falk & Darling-Hammond, 2009), the way in which the Head of Curriculum endeavoured to help teachers understand how such tests could be useful as part of a broader array of data – including daily classroom data – also reveals more educative teacher learning governance processes associated with the deployment of these numbers – including a logic of appropriation of more performative influences (Author, 2014).

The Head of Curriculum's efforts to encourage teachers to use the PM levels, NAPLAN and other standardised tests (PAT tests) to inform their teaching of the curriculum (C2C/KLAs) more broadly resonates with research into how standardised test scores could be used to improve teaching and learning (Chen, Salahuddin, Horsch & Wagner, 2000), at least some of the time (Datnow & Hubbard, 2015). This also potentially encourages the use of more standardized testing primarily for formative and diagnostic purposes (cf. Chisholm and Wildeman's (2013) research in the South African context). Teachers engaged in professional learning discussions as vehicles geared towards improved understanding of the nature of the curriculum they were teaching and the various sorts of data they generated through their work with students. They engaged with a wide array of activities as part of the Inquiry Cycle approach more broadly, and evidence of student learning, including bookwork and assessment samples, and were genuine in their commitment to try to use this time to improve their understanding of the curriculum, and subsequent practice. Also, discussions amongst these teachers, involving shared moderation of assessment against specified standards, were indicative of the learning that occurs through such collective experience (Klenowski, 2011).

As with concerns about benchmark tests in the United States that teachers felt were of little value (Bancroft, 2010), and about which they felt there were inadequate professional development opportunities (Datnow & Hubbard, 2015), teachers at Lillyvale expressed concern about the value and



validity of the PM benchmarks which they collected vis-à-vis everyday classroom practices, and in light of the evidence of the folios of students' learning they had before them with documentation in the form of both standardised and classroom data (cf. Krechevsky *et al.*'s (2009) and Duckor & Perlstein's (2014) advocacy for evidence of students' learning on substantive classroom tasks).

However, such potentialities need to be proffered cautiously, and always in the knowledge that the numbers generated through standardised approaches to capturing student learning, including reading practices, have perhaps unpredictable effects, and sometimes, as evident here, perverse effects, including appearing to actually decrease teachers' data literacy; unpredictability resonates with enactment of policy and efforts to foster professional learning practices at the local level (Watson & Michael, 2016). DeBard and Kubow (2002) highlight how national testing in the US fostered compliance with testing, but not necessarily the level of commitment for learning to ensure substantive improvement in educational practice; in the context of Ohio, this was expressed through 'how current reform agendas have been centered on measurement and have resulted in comparisons among rather than progress within school districts' (p. 390). How such data are actually deployed is crucial to whether or not they are beneficial.

Conclusion

Teachers' learning, as expressed through a particular collaborative inquiry approach at one school site, was governed by a strong focus upon the enumeration of students' learning, expressed through various forms of standardized school, regional, state and national tests/data. Students' attainment in relation to 'levelled' readers, progressive assessment tests, and the results on NAPLAN had a significant impact upon whether and how teachers were construed as 'successful' in their practice, and this in turn played out as key foci as part of their ongoing learning, and manifest as sometimes perverse effects in relation to teachers' data literacy practices. At the same time, the data presented here also reveal that more reductive approaches to standardization of learning can be challenged by teachers – that such rationalities are productive of critique even as they constitute compliance. At the same time as these educators seemed to be dominated by more reductive and technocratic approaches to their own and students' learning, they sought to challenge such processes. These governance processes are multifarious in nature, and potentially productive of a more educative disposition in relation to teachers' learning.

However, the data also reveal that teachers' learnings are always and everywhere actively governed by these standardised measures, and that these governance processes transpire in complex and seemingly contradictory ways. These teachers were always and everywhere *responsive* to the quantification and standardization of education; given the broader global neoliberal policy imaginary (Rizvi & Lingard, 2010) within which NAPLAN was undertaken in Australia, they could not but be so disposed.



Consequently, practically, the research argues that standardised schooling governance processes are sophisticated in their intent and how they become enacted in practice. An important contribution of the research is also to reveal how more perverse effects associated with numeric standardized data can be expressed not only systemically (Lingard & Sellar, 2013), but also in school settings, and in the context of ongoing teacher learning initiatives. And these perverse effects include what appear to be reduced data literacy amongst teachers, as various data discussions (purportedly designed to enhance teacher learning) become reconstituted as sites for the reinforcement of the deployment of abstract indices of teacher quality, disconnected from the circumstances of their production, and actual student learning practices.

In the context of teachers' learning, such processes need to be engaged with cautiously to mitigate the problematic, reductive practices that attend standardisation of educational practice, even as various forms of data have the potential to cultivate more educative dispositions amongst those constituted through their practices.

In this way, the research cautions against the influence of such standardisation governing processes for how they potentially limit teachers' learning to more narrow measures/conceptions of students' learning, even as more collaborative inquiry-oriented teacher learning practices provide the opportunity for teachers to critique these more reductive effects.

References:

Author. 2014.

Author. 2015.

Angus, L. 2015. "School choice: neoliberal education policy and imagined futures." *British Journal of Sociology of Education* 36 (3): 395-413.

Ashadi, A, and Suzanne, R. 2016. "High stakes testing and teacher access to professional opportunities: lessons from Indonesia." *Journal of Education Policy* DOI: 10.1080/02680939.2016.1193901

Au, W. 2011. "Teaching under the new Taylorism: high-stakes testing and the standardization of the 21st century curriculum." *Journal of Curriculum Studies* 43 (1): 25-45.

Ball, S. 1994. *Education Reform: A Critical and Post-structural Approach*. Buckingham: Open University Press.

Ball, S. 2003. "The teacher's soul and the terrors of performativity." *Journal of Education Policy* 18 (2): 215-228.



- Ball, S. 2015. "Editorial: Education, governance and the tyranny of numbers." *Journal of Education Policy* 30 (3): 299-301.
- Bancroft, K. 2010. "Implementing the mandate: The limitations of benchmark tests." *Educational Assessment, Evaluation and Accountability*, 22 (1): 53-72.
- Bergman, B. 2012. "Put data in the driver's seat." *Journal of Staff Development* 33 (4): 44-48.
- Berlant, L. 2011. *Cruel Optimism*. Durham, NC: Duke University Press.
- Bousfield, K., and Ragusa, A. 2014. "A sociological analysis of Australia's NAPLAN and My School Senate Inquiry submissions: the adultification of childhood?" *Critical Studies in Education* 55 (2): 170-185.
- Carter, M., Klenowski, V., and Chalmers, C. 2016. "Who pays for standardised testing? A cost-benefit study of mandated testing in three Queensland secondary schools" *Journal of Education Policy* 31 (3): 330-342
- Chen, J., Salahuddin, R., Horsch, P., and Wagner, S. 2000. "Turning standardized test scores into a tool for improving teaching and learning: An assessed based approach." *Urban Education* 35 (3): 356-384.
- Chisholm, L., and Wildeman, R. 2013. "The politics of testing in South Africa." *Journal of Curriculum Studies* 45 (1): 89-100.
- Comber, B. 2012. "Mandated literacy assessment and the reorganisation of teachers' work: Federal policy, local effects." *Critical Studies in Education* 53 (2): 119-136.
- Datnow, A., and Hubbard, L. 2015. "Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future." *Teachers College Record* 117 (4): 1-26.
- DeBard, R., and Kubow, P. 2002. "From compliance to commitment: The need for constituent discourse in implementing testing policy." *Educational Policy* 16 (3): 387-405.
- Duckor, B., and Perlstein, D. 2014. "Assessing habitus of mind: Teaching to the test at Central Park East Secondary School." *Teachers College Record* 116 (2): 1-33.
- Erskine, J. 2014. "It changes how teachers teach: How testing is corrupting our classrooms and student learning." *Multicultural Education* 21 (2): 38-40.
- Falk, B., and Darling-Hammond, L. 2009. "Documentation and democratic education." *Theory into Practice* 49 (1): 72-81.



Fiore, L., and Suárez, S. (2010). "This issue." *Theory into Practice* 49 (1): 1-4.

Gable, A., and Lingard, B. 2013. "NAPLAN and the performance regime in Australian schooling: A review of the policy context." University of Queensland Social Policy Unit Paper No. 5. Brisbane: University of Queensland.

Glazer, J., and Peurach, D. 2015. "Occupational control in education: The logic and leverage of epistemic communities." *Harvard Educational Review* 85 (2): 172-202.

Grattan, M. 2012. "PM pledge for top five school spot." *The Sydney Morning Herald*. Accessed December 5, 2015. <http://www.smh.com.au/federal-politics/political-news/pm-pledge-for-top-five-school-spot-20120902-258k5.html>

Grek, S. 2008. "From symbols to numbers: The shifting technologies of education governance in Europe." *European Educational Research Journal* 7 (2): 208-218.

Grek, S. 2009. "Governing by numbers: The PISA effect in Europe." *Journal of Education Policy* 24 (1): 23-37.

Hutt, E. 2014. "The GED and the rise of contextless accountability." *Teachers College Record* 116 (9): 1-20.

Klenowski, V. 2011. "Assessment for learning in the accountability era: Queensland, Australia." *Studies in Educational Evaluation* 37 (1): 78-83.

Klenowski, V. 2016. "Questioning the validity of the multiple uses of NAPLAN data." *National testing in schools: An Australian Assessment*, edited by B. Lingard, G. Thomson, and S. Sellar, 44-56. London: Routledge.

Klenowski, V., and Wyatt-Smith, C. 2012. "The impact of high stakes testing: The Australian story." *Assessment in Education: Principles, Policy & Practice* 49 (1): 65-79.

Krechevsky, M., Rivard, M., and Burton, F. 2009. "Accountability in three realms: Making learning visible inside and outside the classroom." *Theory into Practice* 49 (1): 64-71.

Labaree, D. 2011. "The lure of statistics for educational researchers." *Educational Theory* 61 (6): 621-632

Labaree, D. 2014. "Let's measure what no one teaches: PISA, NCLB, and the shrinking aims of education." *Teachers College Record* 116 (9): 1-14.



- Lingard, B. 2010. "Policy borrowing, policy learning: testing times in Australian schooling." *Critical Studies in Education* 51 (2): 129-147.
- Lingard, B., Martino, W., Reazai-Rashti, G., and Sellar, S. 2016. *Globalizing Educational Accountabilities*. New York: Routledge.
- Lingard, B., and Rawolle, S. 2011. "New scalar politics: Implications for education policy." *Comparative Education* 47 (4): 489-502.
- Lingard, B., & Sellar, S. (2013). 'Catalyst data': perverse systemic effects of audit and accountability in Australian schooling. *Journal of Education Policy* 28(5): 634-656.
- Mandinach, E., and Gummer, E. 2015. "Data-driven decision-making: Components on the enculturation of data use in education." *Teachers College Record* 117 (4): 1-8.
- Mehta, J. 2015. *The Allure of Order: High Hopes, Dashed Expectations and the Troubled Quest to Remake American Schooling*. Oxford: Oxford University Press.
- Meyer, H-D., and Benavot, A. 2013. *PISA, Power and Policy: The Emergence of Global Educational Governance*. Oxford: Symposium Books.
- Miller, P., and Rose, N. 1990. "Governing economic life." *Economy and Society* 19 (1): 1-31.
- Mockler, N. 2013. "Teacher professional learning in a neoliberal age: Audit, professionalism and identity." *Australian Journal of Teacher Education* 38 (10): 35-47.
- Mockler, N. 2015. "From surveillance to formation? A generative approach to teachers 'performance and development' in Australian schools." *Australian Journal of Teacher Education* 40 (9): 117-131.
- Moore, A., and Clark, M. 2016. "'Cruel optimism': teacher attachment to professionalism in an era of performativity" *Journal of Education Policy* 31 (5): 666-677.
- O'Leary, M., and Wood, P. 2016. "Performance over professional learning and the complexity puzzle: lesson observation in England's further education sector." *Professional Development in Education*, DOI: 10.1080/19415257.2016.1210665
- Ozga, J. 2008. "Governing knowledge: Research steering and research quality." *European Journal of Educational Research* 7 (3): 261-272.
- Ozga, J. 2009. "Governing education through data in England: From regulation to self-evaluation." *Journal of Education Policy* 24 (2): 149-162.



- Pease-Alvarez, L., and Thompson, A. 2014. "Teachers working together to resist and remake educational policy in contexts of standardization." *Language Policy* 13: 165-181.
- Piro, J., and Mullen, L. 2013. "Outputs as educator effectiveness in the United States: Shifting towards political accountability." *International Journal of Educational Leadership Preparation* 8 (2): 59-77.
- Polesel, P., Rice, S., and Dulfer, N. 2014. "The impact of high-stakes testing on curriculum and pedagogy: a teacher perspective from Australia." *Journal of Education Policy* 29 (5): 640-657.
- Porter, T. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Ready, D. 2013. "Associations between student achievement and student learning: Implications for value-added school accountability models." *Educational Policy* 27 (1): 92-120.
- Rizvi, F., and Lingard, B. 2010. *Globalizing Education Policy*. New York: Routledge.
- Rose, N. 1991. "Governing by numbers: Figuring out democracy." *Accounting, Organizations and Society* 16 (7): 673-692.
- Shank, G. 2006. *Qualitative Research: A Personal Skills Approach*. Upper Saddle River, NJ: Pearson.
- Simola, H., Ozga, J., Segerholm, C., Varjo, J., and Andersen, V. 2011. "Governing by numbers: The rise of data in education." *Fabricating Quality in Education: Data and Governance in Europe*, edited by J. Ozga, P. Dahler-Larsen, C. Segerholm, and H. Simola, 96-106. London: Routledge.
- Sellar, S., & Lingard, B. 2013. "The OECD and global governance in education." *Journal of Education Policy* 28 (5): 710-725.
- Thompson, G., and Mockler, N. 2016. "Principals of audit: testing, data and 'implicated advocacy'." *Journal of Educational Administration and History* 48 (1): 1-18.
- Thompson, G., and Cook, I. 2014. "Manipulating the data: teaching and NAPLAN in the control society." *Discourse: Studies in the Cultural Politics of Education* 35 (1): 129-142.
- Timperley, H. 2009. "Using assessment data for improving teaching practice." Paper presented at the Australian Council for Educational Research Conference. Accessed October 4, 2015. http://research.acer.edu.au/research_conference/RC2009/17august/7/



Twombly, S. 2014. "When teaching interferes with learning: Balancing accountability with the unique needs of every child." *The New Educator* 10 (1): 44-52.

Van Leent, L, and Exley, B. 2013. "Literacy coaching roles in diverse contexts of teaching and learning: New ways of working." *Australian Journal of Teacher Education* 38 (4): 17-30.

Watson, C., and Michael, M. 2016. "Translations of policy and shifting demands of teacher professionalism: from CPD to professional learning" *Journal of Education Policy* 31 (3): 259-274.

Williamson, B. 2015. "Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments." *Journal of Education Policy* 31 (2): 123-141.

Wiseman, A. 2013. "Policy responses to PISA in comparative perspective." In *PISA, Power and Policy: The Emergence of Global Educational Governance*, edited by H-D Meyer, and A. Benavot, 303-322. Oxford: Symposium Books.

Yin, R. 2016. *Qualitative Research from Start to Finish*. 2nd ed. New York: The Guilford Press.

Dr Ian Hardy is Senior Lecturer, and Australian Research Council (ARC) Future Fellow at the School of Education, The University of Queensland. Dr Hardy researches and teaches in the areas of educational policy and politics, with a particular focus upon the nature of teachers' work and learning. As an ARC Future Fellow (2015-2018), Dr Hardy is currently undertaking full-time research into how policy support for curricula reform influences teacher learning in Queensland, within a broader global policy context. At the same time, Dr Hardy is exploring how concurrent policy reform in Scandinavian (Finland and Sweden) and North American (Ontario and Connecticut) contexts is currently constituted, and influencing practice.

This is an earlier version of a paper which was later published as: Ian Hardy (2018) Governing teacher learning: understanding teachers' compliance with and critique of standardization, *Journal of Education Policy*, 33:1, 1-22, DOI:10.1080/02680939.2017.1325517. The peer reviewed paper can be found at <https://www.tandfonline.com/doi/full/10.1080/02680939.2017.1325517>

Dr Hardy can be contacted at i.hardy@uq.edu.au and + 61-7-3365 6425