doi: http://dx.doi.org/10.7203/RASE.10.2.10050
Págs. 192-208

Measuring, Monitoring, and Managing for Productive Learning? Australian Insights into the Enumeration of Education

¿Medición, seguimiento y gestión al servicio del aprendizaje productivo? Enseñanzas australianas respecto a la enumeración de la educación

Ian Hardy[1]

Abstract

This article reveals how ongoing teacher learning in schools is heavily influenced by the increased measurement of student learning outcomes. It draws upon a broad range of literature on the practice of education, measuring learning, and its effects on ongoing teachers' learning at the school site. The paper analyses the meeting transcripts of an ongoing, long-term teacher learning initiative in a school in northern Queensland, Australia, and indicates how processes of measurement influenced such learning. Teachers' learning is revealed as heavily influenced by processes of «measuring [student] learning», the subsequent «monitoring» of student learning through students' data profiles, and 'managing' the learning of lower performing students. The article cautions that while these resultant teacher learning practices may be beneficial for students' learning, this is a contested point, and requiring much more careful analysis of the effects of the measurement of education on such practices.

Keywords

Quantification of learning; testing, standardized testing, standardized learning, teacher learning, teacher professional development.

Resumen

Este artículo muestra cómo la formación continua de los maestros en las escuelas está muy influida por el incremento de la medición de los resultados del aprendizaje de los estudiantes. Se basa en un amplio repertorio de literatura sobre la práctica educativa, la medición del aprendizaje y sus efectos sobre la formación continua de los docentes en la escuela. El texto analiza las transcripciones de una serie de reuniones en el marco de una iniciativa de formación continua a largo plazo de docentes en una escuela en el norte de Queensland, Australia, e indica cómo los procesos de medición influyeron en esa formación. Se constata que la formación de los maestros está muy influida por los procesos de «medición del aprendizaje de los estudiantes», el seguimiento posterior del aprendizaje estudiantil través de sus perfiles de datos y la «gestión» del aprendizaje de los estudiantes de bajo rendimiento. El artículo advierte que, aunque estas prácticas resultantes de la formación del profesorado pueden ser beneficiosas para el aprendizaje del estudiantado, esta es una cuestión a debate, y requiere un análisis mucho más cuidadoso de los efectos que la medición de la educación ejerce sobre tales prácticas.

Palabras clave

Cuantificación del aprendizaje; pruebas; pruebas estandarizadas; aprendizaje estandarizado; formación del maestro; desarrollo profesional docente.

Recibido: 31-03-2017
Aceptado: 28-04-2017

1 University of Queensland, (Australia); ian.hardy@uq.edu.au

1. Introduction: Learning in an era of accountability

This article explores how teacher learning is construed under current conditions of increased accountability for educational outcomes. These processes are part of a broader push away from focusing upon educational attainment as a marker of educational quality within the educational system, to measuring outcomes of schooling practices. As Carnoy argues, «[t]he new emphasis on educational quality has been accompanied and promoted by the rapid spread of testing and measurement» (Carnoy, 2016:34). Measuring and comparing school outcomes have taken on increasing significance in nation-states, as educational policy-makers have sought to deploy such practices to ascertain the «success» of their respective educational systems. These processes have not occurred in isolation but have been fostered through the «soft power» (Nye, 2004) of the International Association for Evaluation of Educational Attainment (IEA), the Organization for Economic Co-operation and Development (OECD), the World Bank, Asian Development Bank, and Inter-American Bank. Such bodies promote a conception of education as equivalent to economic competitiveness, and of improved educational outcomes as key to increased economic productivity. Such pressures are felt at the nation-state level in the push to ensure some way of tracking, of keeping account, of these outcomes over time. This includes monitoring progress on international tests of achievement, such as the IEA's Trends in International Mathematics and Science Survey (TIMSS), and the OECD's Programme for International Student Assessment (PISA).

These processes of keeping account of educational outcomes are also expressed at the national level through national standardized tests of achievement, most notably in the areas of literacy, mathematics and science. In the Australian context, this relates to the national literacy and numeracy tests orchestrated through the National Assessment Program – Literacy and Numeracy (NAPLAN).

Within this context, it is important to ascertain how such broader processes and foci influence the nature of the learning supported within schools. This includes teachers' learning. In this paper, I draw upon transcripts of professional development meetings of teachers in one school in north Queensland, Australia, to reveal how measuring learning outcomes has influenced what is considered important teacher and student learning. The meetings, described as «Inquiry Cycles», were day-long events involving teachers at each year level meeting together to inquire into the nature of their students' learning.

2. The challenge of conceptualising education and learning: «Practical» accountability

To ascertain how the measurement of learning is expressed in schooling settings, it is firstly necessary to consider the purposes of education more broadly. Education is an inherently «practical» activity, insofar as it is grounded in the everyday practices of teachers and students. Drawing upon Aristotelian theorising, Biesta (2013) frames this as very much a problem of how to respond to ongoing, practical problems. In trying to understand what education is, Biesta (2013) calls us to consider a number of binaries about the nature of the practice of education. If we are to take up the call to consider education as an open system of meaning (rather than «a causal, deterministic system of inputs and outcomes»), as focused upon interpretive encounter (rather than «an anonymous machine»), and as «event-like» (rather than «production-like») in its intent, then we are called to enact a conception of education as one which is premised upon particular decisions at particular moments (Biesta, 2013:686). Furthermore, these decisions are never straightforward, but always open to consideration, and determined by context; as Schwab (2013/1970) argues, such decisions are inherently «practical». The practical in this instance is always something immediate and grounded; it «is always something taken as *concrete* and *particular*

and treated as indefinitely susceptible to *circumstance*, and therefore highly liable to unexpected *change*» (Schwab, 2013/1970:593; emphasis added). On this rendering of education, accountability pertains to ensuring conditions for the fostering of deliberative encounters with practical problems in consultation and collaboration with others. Understanding education is neither inductive nor deductive, but inherently «deliberative», and focused upon choosing «not the *right* alternative, for there is no such thing, but the *best* one», under the circumstances (Schwab, 2013/1970:618; emphasis original). Such deliberation resists precise calculation and is dependent upon accountability in the form of teacher judgment.

3. Quantifying learning in an era of enumerative accountability

However, there are myriad forms of accountability within schooling systems, and more «practical» conceptions, based on teacher judgment, struggle for recognition within current educational policy discourses of increased measurement of learning. This measurement of learning is a product of the increased reliance upon numbers in education, itself a manifestation of perceptions of a relative lack of status, or relatively poor standing within which this call to numbers is rendered. Porter (1995) refers to a «trust in numbers» to try to capture perceptions of the benefit and value of numbers. Numbers are seen to give an air of gravitas, a sense of «objectivity» to the particular practices to which they pertain. Indeed, in more complex environments, characterized by conflicting perspectives about how best to proceed – such as educational settings – more «objective» modes of reasoning are able to exert disproportionate influence; Porter refers to numbers as associated with «mechanical objectivity», and foregrounds how complex environments encourage such mechanical objectivity: «Where a consensus of experts is hard to reach, or where it does not satisfy outsiders, mechanical objectivity comes into its own» (Porter, 1995:4). Mathematical and quantitative accounts are seen as particularly effective in justifying specific positions. This is the case even as such figures are themselves never able to capture the nature of the actual practices to which they pertain. Even as such quantified information is in a sense artificial, such information helps constitute a «reality» which may be difficult to deny. Under these circumstances, «rival measures are… at a great disadvantage» (Porter, 1995:5-6)

The dominance of such measures is also resonant in the ways in which various «state-numbers» or statistics actually help to constitute the phenomena to which they purport to relate; in this sense, such numbers are complicit in the ways in which particular phenomena come to be understood and «known» (Desrosières, 1998). Sætnan, Lomell and Hammer (2011) summarise this succinctly: «society and the statistics that measure and describe it are mutually constructed» (p. 1). Various forms of statistics are both productive of specific and actual circumstances, at the same time as they represent constructs of the worlds to which they relate (Desrosières, 1998:3). Through the act of counting, we do not exist beyond that which we seek to measure but rather «enter into it, redefine it, change the stakes that affect it» (Sætnan, Lomell & Hammer, 2011:1). The power of such measures is a response to efforts to develop more «objective measures» of particular phenomena, but in so doing, such measures construct knowledge in their own image.

4. An infrastructure of accountability

Globally in schooling, the power of numbers is particularly evident via what Lingard, Martino, Rezai-Rashti and Sellar (2016) refer to as forms of «globalizing, top-down, test-based modes of accountability» (p. 14). Such renderings of accountability are part of a broader global data infrastructure that construes particular kinds of measures of student learning, such as OECD PISA results, as more significant than

others. While bodies such as the OECD are not able to exert direct influence over educational practices in sovereign nation-states, through support from member nations for it to undertake its work, the OECD has been able to adopt a position of «soft power» (Nye, 2004) over (and beyond) member states. Through processes of comparison enabled by such numbers, politicians have been able to use such data as evidence of the need for further interventions in education (Novoa & Yariv-Mashal, 2003).

These processes of enumerating and comparing learning are enabled by what Anagnostopoulos, Rutledge and Jacobsen (2013) refer to an «infrastructure of accountability», which simultaneously constitutes and «collects» the data it «records». This infrastructure includes intricate, large-scale information systems, operating to collect data across schools, regions and educational authorities, and purports to be able to collect, analyse and disseminate information about the nature of student learning, and the effects of teachers' teaching practice. In English and European contexts, Lawn (2013) refers to the «rise of data» in education systems to try to capture the nature of these enumerative technologies of control that enable particular kinds of collection, visualisation and use of data. Williamson (2016) also refers to how numbers and associated forms of computational data provided through schooling systems and associated bodies provide important visual representations of practice, constituting the forms of educational governance processes that arise, including through various forms of profiles, summaries and comparisons of performance.

5. «Data use» and Teachers' learning

In the arena of teachers' learning, this quantification of education is often expressed in relation to various forms of «data use». Such data use is evident in efforts to shift the focus from engaging in various forms of professional development oriented towards how to engage with data, to the use of data as a form of professional development (Vanhoof & Schildkamp, 2014). This is often described in literature on teachers' learning as «data-based decision making». Schildkamp and Lai (2013) define data use, or «data-based decision making» as processes whereby 'schools make decisions about students, about instruction, and about school and system functioning based on a broad range of evidence, such as scores on students' assessments and observations of classroom teaching' (p. 1). Relatedly, Lai and Schildkamp (2013) take a broad-based approach to data, which they understand as any «information that is collected and organized to represent some aspect of schools» (p. 10). Such definitions suggest a broad conception of data including, but not limited to, more quantitative conceptions articulated earlier. Vanhoof and Schildkamp (2014) argue effective use of data enables educational processes and fosters enhanced student achievement. However, actually engaging substantively with such data is challenging work; Schildkamp, Karbautzki and Vanhoof's (2014) research into how data was actually used in schools in the UK, Germany, Poland, Lithuania and the Netherlands revealed that even as data were collected for multiple purposes, including for policy development, school development/improvement, teacher professional development, and broader accountability purposes, the actual use of data was a much more superficial process, and underdeveloped.

In the Australian context, Klenowski (2016) questions the validity of the varied uses of NAPLAN data. Reflecting upon the OECD's (2013) emphasis upon the importance of a variety of formative and summative assessment tasks, and O'Neill's (2013) differentiation between primary and secondary uses of test data at school and systemic levels, Klenowski (2016) argues the dominance of secondary uses of data calls into question the validity of such approaches; she questions «whether the policy can deliver accountability without negatively impacting high quality and high-equity teaching and learning» (p. 44).

At the same time, as Groundwater-Smith and Mockler (2009) reveal, even as teachers' work and learning are influenced by more reductive approaches to curriculum, teaching and testing, they simultaneously endeavour to engage in more practice-based approaches, rather than simply «complying» with more reductive accounts. More localized conditions and curricula are not simply ignored, even as globalizing and transnational policy processes exert influence (Sivesind & Wahlström, 2016). And testing practices and outcomes can be used for more diagnostic purposes, for determining areas of need, and for improvement, rather than for more performative accountability purposes (Hamilton *et al.*, 2013).

6. Contextualising testing: The Australian experience

In the Australian context, the enumeration of education is most obvious in the form of the National Assessment Program – Literacy and Numeracy (NAPLAN), a national testing regime administered by Australian Curriculum, Assessment and Reporting Authority (ACARA), a national authority endorsed by the federal and state Ministers of Education. State and territory ministers argued for the development of national tests in 2007, and the subsequently elected Rudd Labor Government took up this call (Lingard, Thompson & Sellar, 2016). Interestingly, national testing preceded the development of a national curriculum in Australia, and it is only during 2017 that it is expected that the curriculum (and other reforms) will better articulate with national testing.

The NAPLAN testing regime comprises a census-style test, undertaken every year by all students throughout the country in Years 3, 5, 7 and 9. Feedback is provided to parents and students as an individual report which indicates which «band» or level (out of a 9 point scale) students attained. This is further complemented by a report delivered at the school level on students' performance. This includes patterns in the item-by-item responses, including in comparison with «like» (similar) schools. Perhaps the most overt form of monitoring and visualisation of national testing data occurs through the display of such data via a public website to enable public scrutiny of student results, at the individual school level. The *MySchool* website was introduced in 2010 on the assumption it would provide enhanced transparency into school outcomes, read as literacy and numeracy outcomes on the NAPLAN test. The website compares school performance against 60 statistically similar schools throughout Australia, and national average performance. The «like» schools are denoted through the ICSEA index – The Index of Community Socio-educational Advantage – developed by ACARA as a means of acknowledging the effects of socio-economic status on student performance.

7. Conceptualising learning in context: Penn Primary and the «Inquiry Cycles»

To help understand the nature of these processes in context, and how enumeration processes play out in practice, the research draws on recently collected data drawn from an ongoing 4-year study into the nature of teacher learning practices at one school in north Queensland, Australia. The school serves a lower to middle socio-economic community in a regional city in the northern half of the state. Students come from a range of backgrounds in which both parents work, some parents work, or where parents may be unemployed. Some students live in single-parent families. The school has approximately 850 students, with roughly equal numbers of boys and girls, and 10% of students are Indigenous.

The data comprise a series of meeting transcripts of teachers deliberating together about the nature of the curriculum they were teaching, and the effectiveness of their teaching, in light of evidence of student work. These meetings, known as «Inquiry Cycle», were stimulated by the work of literacy educa-

tor Helen Timperley (2011), who advocated teachers engage in a cycle of investigation into their work, with an emphasis upon evidence of student learning. A key focus of attention during these meetings were various forms of student data collected over the course of students' schooling. This included data in the form of student work in response to formative and summative assessment tasks, students' book-work, NAPLAN data, as well as other forms of data collected at the regional and school level. Regional data included standardized literacy and numeracy data collected through standardized «Progressive Assessment Test – Reading (PAT-R)», and «Progressive Assessment Test – Mathematics (PAT-M)» which students sat at the beginning and end of each year (with the difference between the two results purportedly reflecting students' learning over that year). At the same time, at the school level, teachers were also required to collect various forms of «levelled» reading data, known as «PM Benchmark data»; these data were generated through students' engagement with a set of standardized reading books (graded from 1 to 30), and developed to ascertain students' reading fluency and comprehension capabilities.

8. The data

The data comprise the most recent set of Inquiry Cycle meetings collected up to this point, in July 2016. This set of meetings was selected to provide detailed insights into the nature of teachers' deliberations about data, in the context of a strong push for data-based decision making in the Queensland context. As part of a process of «combating threats to validity» (Maxwell, 2013), these data are also recognized as part of a larger corpus of data of 10 sets of meetings attended between 2013 and 2016, the collection of which represent ongoing and detailed engagement in the work of the Inquiry Cycles over time.

The data comprised written transcripts of the most recent set of Inquiry Cycle meetings. Meetings involved all year level teachers, alongside senior members of staff (deputy principal), and specialist staff (e.g. special needs teachers), an external critical friend, and the researcher (author). All names reported are pseudonyms. Each year level met for one day per term to discuss the nature of the curriculum they were teaching their students, and to develop a better understanding of the curriculum. The meetings also entailed teachers analysing evidence of their students' learning – work samples from students' books, samples of work from assessment tasks, as well as standardized test results. Each meeting was recorded, and transcribed remotely.

Within these data, various forms of numeric measures were most explicitly expressed within the Year 2, Year 4 and Year 5 Inquiry Cycles. These data were purposively selected (Yin, 2015) to understand how such numbers and various forms of measurement were deployed during the Inquiry Cycles; these data were analysed in light of literature and theorising into the nature of education, the increased enumeration and datafication of schooling, and teachers' uses of data. The composition of each year level group is outlined below in Table 1.

Table I. Composition of year level groups in which numeric measures were expressed explicitly

Year Level	Year 2	Year 4	Year 5
Composition of Group	Facilitator (deputy-principal; female)	Facilitator (deputy-principal; female)	Facilitator (deputy-principal; female)
	Acting head of Curriculum (female)	Acting head of Curriculum (female)	Acting head of Curriculum (female)
	Critical friend (female)	Critical friend (female)	Critical friend (female)
	6 classroom teachers (all female)	6 classroom teachers (3 male; 3 female)	5 classroom teachers (2 male; 3 female)
	Author (male)	Author (male)	Author (male)
		1 student-teacher (male)	

These teachers were part of a larger teaching workforce of 43 classroom teachers in the school (6 male; 37 female), with the composition outlined in Table 2 below.

Table II. Composition of teachers across all year level groups

Year Level	Prep	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Composition of Group (number of teachers & gender)	7	5	6	6	3	3	5
	females	females	females	females	females	females	females
					3	2	1
					males	males	male

9. Findings: The enumeration of education at Penn Primary

An emergent thematic analysis (Shank, 2002) of the transcripts revealed three key themes associated with the quantification of education at Penn Primary. These were related to various processes of trying to «measure» student learning within individual classes, «monitoring» such learning across class profiles of data, and «managing» the student learning that subsequently ensued, particularly amongst students deemed as under-performing.

9. 1.The primacy of measuring learning at Penn Primary

Processes of «measuring» learning were clearly evident during the meetings in a variety of ways. This included how teachers were asked by the facilitator of the meeting («Felicity») to indicate the percentage of students who were not yet achieving a pass grade in their year level:

> *«Paddy: Alright I'll go, I've got 20% …*
>
> *Lydia: Okay I have 36% …*
>
> *Ted: Yes, percent failing, it's 19% failing. …*
>
> *Lucy: Yeah, I got 20% failing.* (Year 4 Inquiry Cycle)
>
> *Felicity: How many Phil?*
>
> *Phil: 30%, 8 out of 26.*
>
> *Tess: … I'm 8 as well, so I'm 30%»* (Year 5 Inquiry Cycle).

In Year 2, the identification of these numbers was a challenging experience for some teachers, including a first-year teacher who was struggling to engage his class. The significant proportion of students identified as failing in his class provided the stimulus for a discussion about the importance of ensuring that the activities asked of students were reflected against the criteria against which they were to be assessed, and that anomalies were perhaps the result of not giving the students the opportunity to understand the nature of the task:

> *«Chris: So in mine I've got 63% as failing. I guess just obviously what I'm doing is not working in general, they're not getting what I'm trying to teach them, the way I'm going about it obviously is not working.*
>
> *Felicity: … So we've got to be careful, we don't want to mark these kids down because we didn't teach the task well, and this is your first time in Year 4 and probably the first time dealing with a whole year of C2C [new curriculum] and not just English. So we want to have a look at this closely before this*

term, get the tasks right and so the kids who are able to demonstrate a "C" are actually getting one; that's going to shift your data a lot». (Year 4 Inquiry Cycle).

The facilitator's focus on not «want[ing] to mark these kids down», and emphasis upon «teach[ing] the task well» reflects a more educative, «practical», approach to the particular circumstances at play – including a new teacher struggling to make sense of the curriculum. At the same time, this conversation was made possible by beginning with the nature of the marks themselves; the numbers provided a sense of authority against which it was hard to argue (Porter, 1995).

These numbers, including the percentage of students failing, were seen as especially dramatic in Year 2 more broadly, which was seen as a cohort of students achieving at a particularly low level. There was considerable «consistency» across each of the classes in this regard, and with these students' results from the previous year:

«Felicity: So just for all us to think about because we all own all of the data as well, but we would expect to see some consistency in the classes because they came through fairly consistently apparently, as good as – as reliable as the data was. So now we want to have a look – okay so have you done [class] A, what's [class] A, "D" or below?

Jenny: 52[%]…

Felicity: And B?

Linley: … 29%…

Felicity: Okay 29% yep. C?…

Eleanor: I'm C – 50%…

Cindy: E?…

Felicity: So that's going to make yours 50% so it was 10 out of 21…

Felicity: ... who's D?…

Lisetta: I got 57% D and below…

T: E?

Cindy: Me? … 29%…

Felicity: And then F is?

Jenny [for absent teacher]: 45% failing». (Year 2 Inquiry Cycle).

The power of these numbers was evident in the way in which these figures were considered particularly stark for the cohort as a whole, and evidenced in the way the facilitator began this account with a collective comment about all teachers needing to share the results of the cohort as a whole («we all own all of the data»). Mention of the expectation of «consistency» of high proportions of lower performing students between classes also reflects the power of these numbers in the construction of these students as particularly problematic.

The primacy, but also complexity, of measurement was also evident in the way standardized forms of data were privileged, and other forms of data – including numeric data seen as dependent upon teacher judgment alone – critiqued. What seemed to be valued most were those forms of more standardized quantitative data that seemed to provide «hard evidence» of student achievement – particularly NAPLAN and PAT results; other forms of data, generated by teachers, including the numeric PM Benchmark data, were seen as less reliable – as subject to «just teacher judgement»:

> *«Felicity: But what's written in here really could be anything, other than the NAPLAN, all the rest is just teacher judgement – maybe not the PAT-R. But knowing that, my point there is knowing that Lauretta is a 24 – what does that mean? You know, it doesn't really tell me anything, so getting into her data [daily bookwork, formative and summative classroom assessment] will tell you more and then you can talk to it more».* (Year 4 Inquiry Cycle).

Reflecting the power of NAPLAN in the Australian context, this data was accorded primacy; while the validity of the use of NAPLAN has been questioned more broadly in the Australian context (Klenowski, 2016), in these discussions, there was a sense in which NAPLAN data was ascribed an authority to which other forms of data could only aspire.

Even as there were concerns about some forms of data (including PM numeric data), there was still a valuing of more standardized, numeric forms of evidence. This was expressed in different ways, including in the expectation that teachers would undertake PM Benchmark reading tests with students (using a «running record») if students' results were deemed too low on PAT-Reading tests (e.g. below 5 (half way) on a 10-point («stanine») scale):

> *«Felicity: Here as well, the stanine 4 – anyone stanine 4 and below should have a PM in there for Term 1 and Term 2.»* (Year 5 Inquiry Cycle).

Teachers' accounts of their students' performances were heavily peppered with reference to numeric measures of achievement – whether in relation to PMs, PAT results or NAPLAN – even as they recognized the particular, specific circumstances of students in relation to these measures, such as failing to attend school consistently:

> *«Felicity: I noticed in your's, … is Lesley Bruce, … what's her PM level?*
>
> *Ted: 19-ish, 19…*
>
> *Felicity: … if you look at Joel Coolamon … his NAPLAN's pretty average…*
>
> *Lucy: Lou Smith … like her NAPLAN is not too bad at all …She wasn't too bad on NAPLAN. Lou's big issue is her attendance, I think it's really impacting on her grades…*
>
> *Felicity: That makes sense, … she's a little low on NAPLAN but not disastrous, the same with the PMs and the PAT-R; but, yeah, attendance – okay.»* (Year 4 Inquiry Cycle).

In this way, specific numeric, and often standardized measures of attainment were foregrounded, even as teachers were simultaneously aware of the limitations of these numbers, and of the personal circumstances which influenced their students' learning.

9. 2.Monitoring learning through class profiles

From the outset of the Inquiry Cycle meetings, it was clear that teachers were to be engaged in using these measures to monitor learning using various summaries/class profiles of data at Penn. This was an active undertaking in which teachers were asked to consider the nature of their students' results as a whole, and to identify lower performing students whom they believed could improve, and attain a passing grade for the year. This involved scrutiny of individual teachers' summaries of student data (LoA data, PMs, PAT results and NAPLAN) in their «class profile»:

> *«Felicity: I want to just get into a process to look at your class profile and say, "Okay; who are the kids that we're going to try and shift?"* (Year 4 Inquiry Cycle).

> *Felicity: Over in the data [in class profile] … look at that as well because that's telling a whole other story. But also look at this and think, yeah, but which kids, with just a bit of tweaking are going to get there; because the more that you can catch up to year level this year, the more the [Year] 3s can catch up, after that.»* (Year 2 Inquiry Cycle).

This monitoring, through the visualisation of data (Williamson, 2016) via the class profiles, was also evident in the way teachers compared results across different data sets, drawing upon more standardized and teacher judgment measures to bolster arguments about coherence between the different data sets:

> *«Frances: I'm just looking now, and … I'm matching [Levels of Achievement data] across to the PAT-R and-* (Year 5 Inquiry Cycle).

> *Ted: She's the main one that where the data didn't really line up, the rest of them kind of, like if you look at Liza up the top – the [PAT-R] stanine reflects her LOA [Level of Achievement A to E] data, reflects the PM, and the NAPLAN are pretty much lined up.*

> *Felicity: Yeah it does.»* (Year 4 Inquiry Cycle).

Reflecting the co-constitutive nature of the numbers and society (Sætnan, Lomell & Hammer, 2011) – in this case, in relation to students whose results deemed they did not require further scrutiny – there was a sense in which it was not necessary to closely monitor those students whose «indicators» of learning appeared to be in keeping with one another on the class profile documents, and who were deemed achieving at an «adequate» level.

However, this monitoring of students' learning was not straightforward, and involved close scrutiny of an array of data collected about individual students, and sometimes disparate results within individual students' profiles. Teachers were called upon to justify specific students' outcomes across this array of data, and to explain why there may be anomalies between different data sets within the profiles:

> *«Felicity: So what we're going to do first is get "down and dirty" and interrogate the data, so I've printed your class profiles off, and that, first of all, we'll look at.* (Year 4 Inquiry Cycle).

> *Felicity: I'll just fire questions when I see something that's going to impact on you – and I can see one here, Toby Jones. He's got a [PAT-R] stanine 1 but you think, are you confident that he's a [PM] level 25?*

Tess: Yeah, yeah-

Felicity: Why is he a D then?

Tess: I don't know what happened in that stanine 1.» (Year 5 Inquiry Cycle).

Quantitative measures of data were referred to explicitly in an effort to try to justify why students were achieving as they were according to Level of Achievement data, Pat-R data (stanine 1), and PM Benchmark data (level 25). Monitoring *across* these data sets was construed as a useful means of ascertaining anomalies in terms of specific students' performance. However, this also involved questioning the extent to which certain measures were the «best tool to use»:

«Felicity: So what's her strengths according to this set of data?

Dulcie: She can read but she doesn't understand what she's reading. When I listen to her PMs, she's gone from a 26 to a 25, so when I listen to her read …, so I've just listened to her last term read [PM] 25, she's got 96<% accuracy but her comprehension was only 2 out of 8, so she should even be lower.

Felicity: So how is she getting a [PAT-R] stanine 4 at the beginning of the year?

…

Dulcie: Good guessing. … So I think that might just be a stroke of luck.

Felicity: So, well what about the NAPLAN then? Because her reading was okay and her PAT-R is suggesting that she – it's not brilliant – but it's suggesting that she is reading okay and she's got some comprehension.» (Year 4 Inquiry Cycle).

Such dialogue reveals the Inquiry Cycle process as a vehicle for the simultaneous critique and affirmation of standardized data, and enumerative measures of attainment. Even as there were concerns expressed about the extent to which specific standardized measures were valid (PMs), other forms of data were promoted as potentially helpful for shedding light upon the nature of evidence of actual student learning (NAPLAN); such expression provides evidence of the co-constitutive nature of various forms of numbers and the nature of the teacher learning to which they related (cf. Sætnan, Lomell & Hammer, 2011).

9. 3.Managing learning: «Shifting» and engaging lower performing students

The enumeration processes that underpinned the measuring and monitoring of student learning subsequently leant themselves to the «management» of students' learning, particularly in relation to lower performing students. As alluded to above, such management was evident in efforts to not only monitor students who were currently achieving at a D level or below, but also how to «shift» these students to ensure a greater level of success into the future. This was an «active» process, and seen as only possible in light of the collection (measurement) and scrutiny (monitoring) of student data:

«Felicity: So we'll go through this process of looking at your data, working out which kids are failing at this point in time, because a lot of those won't be getting 'D's by the end of the year, hopefully; that's our goal. It's not to make a judgement on where they're at now, but it's to look at what they currently know and what we need to do next *to shift these kids to a "C".»* (Year 4 Inquiry Cycle; emphasis added).

In this way, there was a proactive process operating within the school that sought to foster teacher learning for student engagement and improvement, and monitoring of various forms of student data were construed as central to this work. As part of this dialogue about data, there was a continued focus upon students deemed likely to perform below year level benchmarks:

> *«Ted: There's a few of those ones that are 'C's though; they're the ones that I focused on in the data conversations [held once per term, with a senior member of staff] that could slip back to 'D's.»* (Year 4 Inquiry Cycle).

Where there were gaps in the data, there was an expectation that teachers would address these by collecting the relevant information:

> *«Felicity: So check that everything's there that should be. Just put a circle around anything that you haven't put in and attend to it as soon as you can.»* (Year 5 Inquiry Cycle).

Managing learning also entailed moving beyond reductive conceptions of data, and foregrounding opportunities to engage more proactively with efforts to foster student learning, and progress, and to recognize students' individual gains:

> *«Ted: And that's where Jessica fell down, I gave her a "D" … [I gave her] a lot of the structure of explaining the language features and what they show you. There was a lot of writing frames involved, and things like that to help her out.*
>
> *Felicity: Yeah and that's great, that's great to give her that success and that's exactly what she needs.* (Year 4 Inquiry Cycle).
>
> *Felicity: More importantly and you guys know this … more importantly than whether they're getting a "D" or not is – are they progressing?*
>
> *Jenny: That's what I – I don't like to get bogged down with, "Oh, is it a 'C' or a 'D'?" I just want to look at individual progress and I know that those children, even though they're a "D", they have made individual gains; and I just don't like having those targets.»* (Year 2 Inquiry Cycle).

The focus on «individual progress» and «individual gains» reflects much more contextualised, «practical» (Biesta, 2013) conceptions of practice, focused upon the peculiar needs of particular students, and how to address these. The targets – in this case, the proportion of students achieving above a "D" on Level of Achievement (summative curriculum) results – were seen as restrictive of the broader notions of learning that were seen as evidence of real progress on the part of students.

This management of student learning also entailed teachers considering the nature of the pedagogies they should enact to help some students remain focused upon their work, and not under-perform. This was evident in the discussion in the Year 5 Inquiry Cycle in relation to one student whose data was disparate, but who was recognized as capable of better learning (and improved outcomes):

> *«Felicity: Look at Mitch Speith. Stanine 2 at the beginning of the year. Apparently he's reading at a level 30; I'd question his comprehension on a whole lot of things for other reasons*
>
> *Tess: But he got stanine 5 last year*
>
> *Felicity: But he got a D …*

Felicity: Right now he's getting a D. He's capable, competent. Why is he getting a D

Tess: Yep

Felicity: Not because he's lazy

Tess: No – he just 'checks out' [becomes distracted in class]

Felicity: We've got to look at … well then – what pedagogies might keep him "in" class, okay? That's the kind of thing.» (Year 5 Inquiry Cycle).

Such responses reflect the multifaceted nature of engagement with and through these numbers, particularly in relation to lower performing students. Critiques of the accuracy of standardized measures of learning, and emphases upon specific pedagogies for specific students, resonate with calls for more practical applications and understandings of education (Biesta, 2013); the very creation of such numbers, potentially, had the capacity to generate conversations and actions about what to do next for particular students in particular circumstances. However, that such conversations as vehicles for teachers' learning were always framed within this paradigm of particular measures of student learning simultaneously frustrates the scope of such potentiality. Better delineating the nature of this quandary requires further inquiry.

10. Discussion and conclusion: Measuring, monitoring, and managing for productive learning?

In many ways, processes of measuring, monitoring and managing educational practices via particular forms of standardized numbers represent forms of «fast policy» (Peck & Theodore, 2015) that have come to characterise processes of statecraft under neoliberal conditions. In educational settings, such processes are associated with not only the enumeration of education, but as evident in the data presented here, processes of constant comparison (Novoa & Yariv-Mashal, 2003) with such measures to «validate» forms of learning. It is the dominance of such measures and monitoring which give pause for thought in relation to the potentially educative effects of the uses of enumerative data in schooling settings, including for fostering teachers' learning. Learning at Penn was conceptualised in light of multiple measures of learning, and the subsequent monitoring and managing practices that these measures enabled. It would not have been possible to engage in the sorts of conversations that characterized the Inquiry Cycles in these year levels if it had not been for the generation of these data, and the various «infrastructure of accountability» (Anagnostopous *et al.*, 2013), including the class profiles, that enabled this process.

Arguably, the engagement with data through the Inquiry Cycle conversations helped to cultivate an understanding of the nature of these data, and how they might inform teachers' learning for enhanced student learning. The use of numbers to identify students who were underperforming, reveals the power of numbers; this was evident in the way in which various measures were foregrounded, particularly the percentage of students deemed underperforming in their Level of Achievement (A to E) data. Such measurement served as a precursor to closer identification and subsequent scrutiny/monitoring of students deemed as requiring interventions to enhance their learning outcomes, and how to best orchestrate learning opportunities for these students. These numbers could be deployed for productive purposes, as a starting point for more detailed cultivation of student and teacher learning. In part, this was the case in relation to the Inquiry Cycles more broadly. It could be argued, therefore, that these

teachers were not simply constrained by broader conditions, within what has been termed an «age of compliance» (Groundwater-Smith & Mockler, 2009), with subsequently reductive effects upon their learning; after all, these teachers were engaged in conversations in the Inquiry Cycle process that did enable more «practical» (Schwab, 2013), practice-based approaches to their learning (Groundwater-Smith & Mockler, 2009). Even as the numbers that peppered their dialogue can be seen as evidence of more globalizing and transnational policy processes, more localized conditions and curriculum were being addressed, and taken into account (Sivesind & Wahlström, 2016). Particular students' needs, in light of the curriculum as taught at the school, were highlighted, and flagged for intervention; such data were being used to try to question the pedagogies in action, and to foster enhanced student achievement (Vanhoof & Schildkamp, 2014), even as this was also challenging (cf. Schildkamp *et al.*, 2014). And there was at least some broad evidence of the use of various PAT and NAPLAN test results for more diagnostic purposes, and not simply for more performative reasons (cf. Hamilton *et al.*, 2013).

However, and at the same time, understandings of student learning were heavily mediated by a focus upon particular kinds of data – data that could be measured, monitored and perceived as useful for «managing» the vicissitudes of learning. That such data existed in so many formats – in this case, NAPLAN, Pat-results, PM Benchmark results, and the proportion of students performing below a pass grade («D») in the regular curriculum – is indicative of how educational quality has indeed been associated with the «rapid spread of testing and measurement» (Carnoy, 2016:34). Significantly, such data were construed as providing important evidence under broader conditions in which teachers' judgment is increasingly under scrutiny. Even as the Inquiry Cycles were explicitly designed to effect enhanced teacher judgment, and involved much discussion about the specific aspects of students' work that needed to be attended to, and interrogation of curriculum to improve teachers' understandings to improve their teaching, the visualisation of enumerative data (Williamson, 2016) and the subsequent focus upon this data within the discussions, also reflects how such judgement alone was considered insufficient, and needed to be complemented by the gravitas that attends numeric representations of student learning; rival «measures» (Porter, 1995) to these numbers were at a distinct disadvantage during these discussions.

As the «pinnacle data» within the Australian national context, NAPLAN data clearly influenced how student learning was construed in the context of the Inquiry Cycles. While individual NAPLAN data were deployed by teachers to provide some insights into the nature of their students' learning, that this work was occurring within a broader context of scrutiny of students' results at the national level in light of both national and international data reveals how the dialogue about data was not simply a «local» phenomenon, but also heavily influenced by the broader policy conditions within which it unfolded. Under such conditions, even as teacher judgement was recognized as important at Penn Primary, there was something of a hierarchy in terms of perceived validity, and «objectivity» (Porter, 1995) of the numbers associated with student learning. NAPLAN was perceived as most authoritative, followed by PAT data; these modes of data were perceived as less subject to teacher judgment than the PM Benchmark data, or teachers' qualitative accounts of student progress, and therefore more useful for «monitoring» student learning. This was the case even as these numbers were recognized as inherently limited and limiting, as needing to be subject to scrutiny on the part of teachers, and as not providing solutions to the problems which they constructed. This was the domain of teacher judgment, even as such judgment was simultaneously critiqued and criticized.

The numbers referred to here were clearly productive of teachers' learning, but the cultivation of teacher judgment more broadly was also simultaneously hemmed in by the parameters of the standardized measures with which they were associated. The way in which such numbers/measures are deployed can potentially limit the sorts of context-relevant, «practical» conceptions of education to which we might aspire as educators (Biesta, 2013). Such numbers are complicit in how particular kinds of knowledge become «known» (Desrosières, 1998), and constitutive of the conditions within which they are generated and deployed (Sætnan, Lomell & Hammer, 2011). In the context of the Inquiry Cycle approach, the potential for these numbers to help inform teachers' understandings of their students' capabilities was evident. However, this potential existed alongside a broader set of conditions which foregrounded particular kinds of numbers as valid and meaningful, and teacher judgment as problematic, even as efforts were made to cultivate such judgment. Such tensions reflect the conflicted and contested nature of the enumeration of education, and ongoing school-based teacher learning, and of the challenges which subsequently attend the measurement, monitoring and management of student learning.

References

Anagnostopoulos, Dorothea; Rutledge, Stacey & Jacobsen, Rebecca (2013): "Mapping the information infrastructure of accountability" in Dorothea Anagnostopoulos, Stacey Rutledge & Rebecca Jacobsen. (eds.): *The Infrastructure of Accountability: Data Use and the Transformation of American Education*. Cambridge: Harvard Education Press.

Biesta, Gert (2013): "Knowledge, judgement and the curriculum: on the past, present and future of the idea of the Practical". *Journal of Curriculum Studies*, 45 (5), 684-696,

Carnoy, Martin (2016): "Educational policies in the face of globalization: Whither the nation state?" in Karen Mundy, Andy Green, Bob Lingard, & Antoni Verger (eds.): *The Handbook of Global Education Policy*. Chichester: Wiley-Blackwell.

Desrosières, Alain (1998). *The Politics of Large Numbers: A History of Statistical Reason*. Cambridge: Harvard University Press.

Groundwater-Smith, Susan & Mockler, Nicole (2009). *Teacher Professional Learning in an Age of Compliance: Mind the Gap*. Rotterdam: Springer.

Hamilton, Laura; Schwartz, Heather; Stecher, Brian & Steele, Jennifer. (2013): "Improving accountability through expanded measures of performance". *Journal of Educational Administration*, 51 (4), 453-475.

Klenowski, Val (2016): "Questioning the validity of the multiple uses of NAPLAN data" in Bob Lingard, Greg Thompson, & Sam Sellar (eds.): *National Testing in Schools: An Australian Assessment*. Abingdon: Routledge.

Lai, Mei Kuin & Schildkamp, Kim (2013): "Data-based decision-making: An overview" in Kim Schildkamp, Mei Kuin Lai & Lorna Earl. (eds.): *Data-Based Decision Making in Education: Challenges and Opportunities*. Springer: Dordrecht.

Lawn, Martin (2013): "Introduction: The rise of data in education" in Martin Lawn (ed.): *The Rise of Data in Education Systems*. Oxford: Symposium.

Lingard, Bob; Martino, Wayne; Rezai-Rashti, Goli & Sellar, Sam (2016). *Globalizing Educational Accountabilities*. New York: Routledge.

Lingard, Bob; Thompson, Greg & Sellar, Sam (2016): "National testing from an Australian perspective" in Bob Lingard, Greg Thompson & Sam Sellar (eds.): *National Testing in Schools: An Australian Assessment*. Abingdon: Routledge.

Maxwell, Joseph (2013). *Qualitative Research Design: An Interactive Approach* (3rd ed.). Thousand Oaks: Sage.

Novoa, António & Tali Yariv-Mashal (2003): "Comparative research in education: A mode of governance or a historical journey?". *Comparative Education*, 39 (4), 423-438.

Nye Jr., Joseph S. (2004). *Soft Power: The Means to Success in World Politics*. New York: Public Affairs.

O'Neill, Onora (2013): "Intelligent accountability in education". *Oxford Review of Education*, 39 (1), 4-16.

Peck, Jamie & Theodore, Nik (2015). *Fast Policy: Experimental Statecraft at the Thresholds of Neoliberalism*. Minneapolis: University of Minnesota Press.

Porter, Theodore (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

Sætnan, Ann; Lomell, Heidi & Hammer, Svein (2011): "Introduction: By the very act of counting – The mutual construction of statistics and society" in Ann Sætnan, Heidi Lomell, & Svein Hammer (eds): *The Mutual Construction of Statistics and Society*. New York: Routledge.

Schildkamp, Kim; Karbautzki, Louisa & Vanhoof, Jan (2014): "Exploring data use practices around Europe: Identifying enablers and barriers". *Studies in Educational Evaluation*, 42, 15-24.

Schildkamp, Kim & Lai, Mei Kuin. (2013): "Introduction" in Kim Schildkamp, Mei Kuin Lai & Lorna Earl. (eds.): *Data-Based Decision Making in Education: Challenges and Opportunities*. Springer: Dordrecht.

Shank, Gary (2002). *Qualitative Research: A Personal Skills Approach*. Upper Saddle River: Pearson.

Schwab, Joseph (2013/1970): "The practical: a language for curriculum." *Journal of Curriculum Studies*, 45 (5), 591-621.

Sivesind, Kirsten & Wahlström, Ninni (2016): "Curriculum on the European policy agenda: Global transitions and learning outcomes from transnational and national points of view." *European Educational Research Journal*, 15 (3), 271-278.

Timperley, Helen (2011). *Realizing the Power of Professional Learning*. New York: Open University Press.

Vanhoof, Jan & Schildkamp, Kim (2014): "From 'professional development for data use' to 'data use for professional development'." *Studies in Educational Evaluation*, 42, 1-4.

Williamson, Ben (2016): "Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments". *Journal of Education Policy*, 31 (2), 123-141.

Yin, R. (2015). *Qualitative Research from Start to Finish* (2nd ed.). New York: The Guilford Press.

Nota biográfica

Ian Hardy es *Senior Lecturer* en The University of Queensland (Australia). Ha analizado las condiciones laborales y de aprendizaje de los docentes con colegas de Australia, Estados Unidos y Canadá. La investigación del Dr. Hardy se centra en la relación entre la educación y la sociedad, y su investigación actual explora las respuestas de los educadores a los contextos sociopolíticos más amplios en los que desarrollan su trabajo. Desde 2004 es miembro del consorcio internacional de investigación «Pedagogía, Educación y Praxis», con la participación de investigadores de Australia, Reino Unido, Suecia, Noruega y Finlandia. Ha publicado como autor *The politics of teacher professional development: Policy, research and practice*. New York, USA: Routledge, 2012; y como coautor: *Changing practices, changing education*. Singapore: Springer, 2014.