Assessment, accountability and the literacy curriculum: re-imagining the future in the light of the past

Author: Gemma Moss, University of Bristol. Email: gemma.moss@bristol.ac.uk

Abstract: This paper explores the dynamic to education system reform in England, highlighting the challenges the range of policy tools in use have created for policymakers, practitioners and researchers alike. Taking literacy policy in England as a "telling case", I will argue that far from holding the answer to successful system reform, high stakes testing and the migration of system-level data into the public domain create a series of challenges for politicians that they cannot satisfactorily solve. In turn, policymakers' attempts to manage the problems away are further compounding the difficulties teachers face. As policymakers' commitments to bringing about well-motivated change waver, research and practitioner communities need to work more closely together to successfully re-imagine a purposeful and engaged literacy curriculum for all.

Introduction

In examining the dynamic to number-driven education system reform this article takes as its telling case (Mitchell 1984) the trajectory to literacy policy in England between 1997 and the current day. The use of numerical data to bring about system change in education is all-pervasive, but yet the precise direction such reform takes varies from system to system, depending upon the configuration of stakeholders, the different powers they wield, the institutional arrangements in which those powers are invested and other aspects of the broader social context that may be harder to pin down. How the school system is configured, the extent of social and economic segregation that may determine who goes to which schools, what counts as a good education in different communities, in public discourse and amongst key stakeholders, all play their part in determining what happens as numerical data are mobilised here as opposed to there (Lundahl and Waldow, 2009).   To understand the role that any specific elements play, the focus in this analysis rests very much with the English case. But the intention is also to bring out some of the more general contradictions in using system level numerical data as the main tool to assess what needs fixing in education and to determine in which direction system reform should go (Sahlberg, 2007). In doing so this article draws on a sequence of research studies that were funded by the ESRC between 2002- 2013[i]. This database has been added to through a process of monitoring and tracking the interaction of literacy policy and practice long after these funded research studies had ended.

Conceptualising policy work as a social imaginary

Policy sociologists studying globalisation increasingly refer to the conjunction of policy tools, discourses and actors and the course of action they bring into being as a "social imaginary" (Rizvi, 2006; Lingard, 2010). Literacy policy in this sense is not so much a recipe or a set of directives established in one place, to be followed in another with greater or lesser fidelity. Rather literacy policy creates a plausible narrative about how things go together that sustains the appropriate patterns of action in the real world and the necessary configuration of resources to bring into being that which is in the process of being collectively imagined. A social imaginary thus has real effects. Tension points arise when the gulf between the original acts of imagination and their realisation in

real world settings becomes apparent. In the case of policy this is an almost inevitable result precisely because policy work is dispersed across time and place and involves very different actors.

Describing literacy policy as work in progress, reliant upon a collective act of imagination, draws attention to the heuristics involved. It also moves attention away from the contents of the policy, and their propositional status – is this a true and necessary description of the sequence to learning literacy or an accurate account of how to change behaviours? – to the consequences of believing it to be so.

> "For a theory to become part of the social imaginary, it must evolve into a kind of common understanding that enables us to carry out our everyday social practices. In this way, social imaginary is both factual and normative, 'that is, we have a sense of how things usually go, but this is interwoven with an idea of how they ought to go, of what missteps would invalidate the practice' (Taylor, 2004 p. 24)." (Rizvi, 2006)

"Social imaginaries" stabilise the everyday, providing the necessary parameters to action by rendering them meaningful. In this sense they are a necessary condition for participating collectively in the life around us. This distinguishes the social imaginary from the concept of ideology, and its counterpart false consciousness, a dichotomy which treats as entirely opposed erroneous beliefs externally imposed from above and authentic knowledge developed locally and from below. Social imaginaries are also multiple, rather than singular. No one social imaginary can fully account for the world as it is. If they can gain traction, they can also lose traction, as the disparate elements they bring into association temporarily coalesce or disperse. Speaking of "economic imaginaries" (ways of imagining economic practices), Jessop explains:

> The totality of economic activities is so unstructured and complex that it cannot be an object of effective calculation, management, governance, or guidance. Instead such practices are always oriented to subsets of economic relations (economic systems, subsystems, or ensembles) that have been semiotically and, perhaps organizationally and institutionally, fixed as appropriate objects of intervention. (Jessop, 2009 p345)

Keeping a social imaginary in place requires continual effort to exclude from the frame what any one imaginary cannot account for. With these conceptual tools in mind, I now turn back to the trajectory to literacy policy in England.

Explaining numbers-driven education system reform in England
A focus on the literacy curriculum is particularly helpful in tracking the dynamic of numbers- driven education system reform, precisely because literacy often finds itself at the forefront of such reform processes. The diverse and multidimensional skill-set literacy represents is understood as the main commodity of exchange that schools create. From a political point of view it has currency. Although the English system has a long history of repeated literacy crises in which worries about whether standards are high enough surface and shape political debate (Moss, 2014; Harper, 2015; Soler and Openshaw, 2006), they do not always lead to the same place. This article is concerned with the specifics of what has happened since 1997 to the present day, given the particular configuration of policy tools, policy ideas, bodies of knowledge, and genres of classroom practice (Lefstein, 2008) that could be drawn on to engineer system reform.

For much of the twentieth century, English education ran without regard to numerical data and with very little active monitoring of school level outcomes. The responsibility for planning the curriculum rested mainly with the teacher, with the prospect of examination at 16 only directly influencing what got taught to whom in the later stages of a secondary school career. The introduction of the National Curriculum in the 1988 Education Reform Act changed all that, bringing with it as it did a specification of curriculum content spanning the main stages of education from primary to secondary school and a system of key stage testing with results made public at school level on an annual basis. Much that has happened to the English education system since depends upon this conjunction. But necessary as they are, in and of themselves these elements are not sufficient to explain what happened next. Rather they created a series of possibilities from which other things could flow.

Elected in 1997, the New Labour government invested most in imagining how the combination of curriculum specification, regular testing and reporting in the public domain [that the Education Reform Act had put in place] could be used to drive system change. They placed a particular emphasis on the reform of literacy and numeracy pedagogy in the primary sector in the first years of their administration, launching the National Literacy and Numeracy Strategies in 1998. The Strategies took the basic tools the Education Reform Act provided and strengthened them, developing a much more elaborate reform architecture that drew on a body of ideas from the school improvement and effectiveness movement. Confidence amongst the key protagonists in the efficacy of such a reform programme created the political will to run the Strategies at scale and they became flagship policies, signalling New Labour's intent to fix public sector schooling through investment (Moss, 2009).

The National Literacy and Numeracy Strategies were designed to standardise delivery and re-shape classroom pedagogy, taking into account a range of ideas about what constituted best practice. If all parts of the system performed at the level of the best, then the assumption was that this would lead to higher performance for all pupils, whilst impacting most directly on schools with the lowest levels of achievement, many of them operating in areas facing the highest levels of social disadvantage. The key ingredients intended to lead to such outcomes were:

- An understanding that the closer specification of curriculum content in the National Strategies would function as a list of deliverables for which teachers could be held to account. This becomes apparent in the shift in the language in which the curriculum sequencing is described (Moss, 2002)
- A means of more closely monitoring curriculum delivery to ensure schools fully covered what had been clearly specified as age appropriate curriculum content, in a given time frame and across the system as a whole. (The planning tools associated with the NLS including the Literacy hour, a uniform focus on word, sentence and text level work within each hour and the progressive goals set for the curriculum in each of these sub-areas)
- Testing at regular intervals to check that progress had been made with encouragement to teachers to monitor their pupils' performance against level descriptors designed to keep their progress on track
- Means of checking across the system so that poorer performance here (less progress, lower attainment) could be challenged by the evidence of higher attainment there
- Increasingly sophisticated statistical modelling to break away from the conviction that higher performing schools simply served more advantaged pupils
- Means of re-gearing how the teacher profession worked so that best practice in the school system and the best attested approaches to pedagogy with the most efficacious outcomes could be widely shared and moved from one place to another

- Means of developing knowledge in the central strategy team so that any lacks elsewhere in the system could be swiftly identified and fixed
- Means of disseminating knowledge from the centre to each part of the system, so that ideas about what worked, what needed adapting and what else might be tried could be directed to those places that might benefit most.

Reassurance that the Strategies would translate into a politically successful project rested as much with the performance management tools they embedded as with the programme of teachers' professional learning that the Strategies resourced. At the time the key terms in which this combination of ingredients were described were "pressure and support". The main debate during the Strategies' lifetime rested with whether the balance between these two elements had been calculated right. The numbers in the system maintained the pressure particularly on those doing least well, on the understanding that this in itself would reduplicate effort and act as a spur to reform. At the same time the data would allow the central strategy team to direct attention and resource to where support was needed most.

Getting it right or getting it wrong? The dangerous game of comparison by numbers
The National Literacy Strategy (NLS) certainly had effects. It changed literacy teaching in primary schools in England, making it more uniform. The literacy hour introduced a new pacing to literacy teaching and new forms of pedagogy, such as "guided reading". Teachers' lesson planning became more geared to achieving well-specified goals: these were made visible in learning objectives to be shared with the pupils at the start of a lesson sequence and written into teacher planning. Practice in one school began to look much more like practice in another. Such a high level of change directly resulted from the complex infrastructure the NLS had put in place including locally-based literacy consultants, a range of mechanisms for sharing teaching resources and best practice within the system, and an on-going programme of nationally planned teacher professional development (See Earl, Watson and Katz, 2009 for a full description). This was accompanied by continual oversight and monitoring of pupil level data by the central team, bound into the political cycle of annual decision-making: reporting to Ministers on progress made, assessing any risks to the policy's success and planning for further expenditure in the year ahead. The high level of resourcing that the Strategies required depended upon an equally high level of political commitment. Yet this also brought attendant risks. Not least that the fortunes of the Strategy were now directly tied to the political fortunes of the party that had championed it so directly. In effect, approaches to teaching literacy embedded in the Strategies, their warrants and their comparative success had become a party political cause.

Placing the Strategies on the public stage in this way created a number of risks for the various stakeholders. From a policymaker's point of view, the NLS needed to publicly demonstrate its success as evidence of the wisdom of the investment made. At the outset of the Strategy a clear goal was set of 80% of pupils reaching Level 4 in Literacy at the end of Key Stage 2 by the year 2002. The promise of performance management – that inputs can be endlessly adjusted to provide better outputs – embedded in the architecture of the NLS suggested that this was an eminently achievable goal. Whilst numbers rose in line with predictions all went well, and at first substantial and sustained early success saw test scores at the end of primary school rise quite dramatically between 1997 and 2000. But from that point onwards the rate of improvement stalled just short of reaching the target Ministers had set. (The goal had been reached in reading but not in writing.) As the plateau in results continued, political support for the strategy began to ebb away. Successive Ministers in the Department of Education started to invest their political time and effort in other areas of their brief. In 2004 when the ballot to nominate the subject of enquiry for the House of Commons Education and Skills Committee fell to a Tory backbencher, a political ambush was launched by the simple expediency of choosing as the next topic "Teaching Children to Read". This

provided an opportunity to highlight what hadn't been achieved by the NLS rather than what had. Through the choice of witnesses called it was possible to mobilise opponents to the NLS who advocated other approaches to the teaching of reading, particularly phonics. As the summary statement at the start of the Committee report said:

> The Government tells us that primary school children have never been more proficient readers. It claims this achievement as the outcome of its National Literacy Strategy (now Primary National Strategy), introduced in 1997. Others question the true extent of this success. Moreover, even if Government figures are taken at face value, at age 11, around 20% of children still do not achieve the success in reading (and writing) expected of their age. This figure is unacceptably high. Furthermore, there is a wide variation in the results achieved by schools with apparently similar intakes. This differential achievement suggests that problems do exist, either in the implementation of the Government's strategies or inherently in the methodology it promotes. (House of Commons, 2005, p3)

The House of Commons Select Committee enquiry was not a research-led evaluation of the efficacy of the National Literacy Strategy, conducted on objective terms. (The policy team had commissioned such a review from an external group of researchers based at OISE (Earl et al, 2002).) Rather it fed into and acted as a sounding board for a public discourse of disappointment with education, in which the numbers always seem to show that education has not done well enough. The reasons why such a discourse of disappointment takes hold may well reflect other factors, not least shifts in the economy which make the certainty of obtaining a good job after a good education more tenuous, and when the numbers of low skill, low wage jobs, offering little security of employment and poor working conditions are increasing. This can leave whole communities adrift from much hope for the future, with little chance of education on its own offering much of a way out.   To create a good news story about education is difficult under these circumstances. Such difficulties are compounded by the mismatch in expectations that the use of performance management tools and assessment instruments in education represent.

Performance management promises equal outcomes for equal inputs. In line with this, the NLS in its first phase put a lot of effort into standardising delivery of the literacy curriculum: ensuring that the same elements were taught in the same ways, using principles of "best practice" which were clearly articulated to teachers and resourced. It may well be that the success of the Strategies in their first phase reflect the impact of just such an investment. But education as a field remains resistant to standardising outputs. Children do not enter the school system on a level playing field. Without paying attention to the levels of disadvantage reflected in school catchments, a list of the best schools produces a list in which schools with more socially advantaged pupils dominate, rather than schools which genuinely represent the best in teaching and learning. Ofsted found this out to its cost in its first attempt to identify "best practice" in the system using numerical data alone (Goldstein, 2001). Attempts to remedy this led to the creation of baskets of schools operating in the same circumstances, on the assumption that they could more usefully benchmark themselves against each other. When system data are used to do this, then reassuringly from a performance management point of view, some schools in each basket are always doing much better than others. They then set the standard which all schools in similar circumstances are expected to achieve, if only they worked in the same way. But this may be to over-interpret the distribution in the dataset that statistical tools create, treating as evidence of well-motivated cause and effect what is in reality a calculation of probability – that in any one year some schools will do substantially better than their peers, and some substantially worse, with most clustering around the mean.

Variation in a dataset is anathema to performance management. It shows a system running below par, with issues in quality that need to be addressed. But for education, variation is often the point

of the exercise. Examination and assessment are designed to elicit differences in student performance. Test instruments will be rejected as inadequate if they cannot do that. Such distinctions are vital in determining the sequencing and pacing that drive curriculum design. The assessment tools used to elicit a child's performance are thus intended to create a trajectory of difficulty: they become the means of judging where any particular child stands, in relation to the topic and in relation to their peers, with more or less conceptual grasp of the issues under consideration. Part of the art of teaching is then to adjust in the light of such assessments, exercising judgement by adapting classroom strategies and approaches so that teaching focuses on what will be most efficacious in taking these children from these starting points forward on the best terms. Taking such variation into account is essential to good teaching.

The capacity to exercise such judgement well depends upon being attuned to the specifics of teaching and learning in its local context and situating what happens here in the longer trajectory of which this moment in assessment is part. Standardising progress obscures these specificities. Thus one fundamental misunderstanding built into the NLS was that writing would develop at the same speed, in the same time frame and according to the same milestones as reading.   In fact there is every reason to think such trajectories are different, that reading precedes writing and influences the latter's course. We may lose sight of this if we think of reading and writing as simply encoding sound into script and vice versa, script into sound. But a competence in reading and writing means much more than this: the acquisition of repertoires of language in its written form and the capacity to re-make such repertoires through use. These depend upon deep investments in acquiring the language of written texts through reading widely first. Curriculum lists may hide such necessary differences in pacing. They may be particularly prone to do so if they construct trajectories in children's learning by focusing on the time taken to deliver a particular sequence of instruction rather than the time children require to build a particular competence. The time it takes a teacher to impart x and then y may differ from the time it takes to get to y, already knowing x. The list cannot capture the necessary interaction and dialogue between given and new that creates the trajectory in learning.

Performance management assumes that all schools will take the same route to creating the best outcomes, that the answers are widely known and can be easily passed around from one school to another and even from one education system to another also. The evidence from large-scale reform programmes shows that this is not the case (Earl et al, 2003; Hargreaves, 2002; Sellar, 2015). Particular systems adjust in the light of their own pasts, adapting the same recipes to different effect[ii]. Local circumstances matter in ways that the numerical outputs cannot fully capture. Take the case of a school identified by a task force set up within the National Strategies towards the end of New Labour's term of office as having exceptionally low performance in comparison to other schools operating in similar circumstances.   On enquiry the task force discovered that the school served an estate used by the Social Services for emergency housing for families in crisis. On investigation the school's exceptionally high pupil turnover of 80% per year was the product of this expediency as families in crisis moved onto the estate and then went on elsewhere to more stable accommodation, once that moment of crisis was over. Under these circumstances it is highly unlikely that the record of Key Stage 1 or 2 results would act as an adequate guide to the efficacy of the school's approach to teaching and learning, or even that the standard prescriptions for good teaching that the NLS provided would indeed be the answer to the pedagogical conundrum the school faced in dealing with this state of affairs.

Handling disappointment on the public stage: what politicians do with the narrative in the numbers

The tension points between the urge to standardise in the tools of performance management, and the continuing need to identify variation in the tools of assessment that education employs create

difficulties in the political sphere, almost regardless of what the numbers actually say. New Labour were particularly vulnerable to a narrative of disappointment given their high profile target of 80% of pupils reaching a promised level of performance within a given timeframe. When the first target of 80% wasn't reached, a second more ambitious target of 85% was announced to demonstrate confidence in the policy trajectory nonetheless.   Having set the policy up with the expectation of continuous improvement, the politicians now found themselves faced with a narrative of failure as each annual test round was published. Successive examples of reporting on the outcomes of Key Stage 1 and 2 results, collected from the BBC website and published respectively in 06 and 07, show something of the flavour.

> **Primary schools miss test targets**
> **This year's league tables of England's primary schools confirm the government missed its targets for the results of Sats tests children took in May.**
> Ministers said attainment had never been better, with 79% of 11-year-olds reaching the expected level in English and 76% doing so in maths. But the aim was for 85% to do so. The opposition said progress had stalled. In just over half of schools, results were higher this year than last - but in almost as many they were worse.
> (BBC reporting 06, http://news.bbc.co.uk/1/hi/education/6210720.stm)
>
> **Fall in primary writing standards**
> **Writing standards among seven-year-olds in England have fallen for the second year in a row, the latest figures show.**
> Some 80% of children reached Level 2 in school writing tests - a fall of one percentage point on 2006. There was no change in maths, science or reading this year, with some 90% achieving the required level in maths, 89% in science and 84% in reading.
> Schools minister Andrew Adonis said he was pleased with the results but called for more efforts to drive up standards. (BBC reporting 07,
> http://news.bbc.co.uk/1/hi/education/6970206.stm)

Evidence for the sustained change in pupil performance that the investment in the Strategies had brought about had been drowned out by reports of missed targets and progress stalled, with any numbers that could be used to make the case for failure attracting most media attention.

The early success and later plateau in the pupil performance outcomes from the National Literacy Strategy certainly raise a number of interesting research questions: about the efficacy of large-scale education reform programmes (Linn, 2002), whether they are helped or hindered by the political support they depend upon (Coburn, 2006), the capacity of any education system to significantly narrow gaps whose antecedents may rest as much with the economy and social segregation as they do with education (Lupton, 2011), and the role of standard versus bespoke solutions in bringing about lasting change.   Clarifying the various factors that helped or hindered the Strategies in doing what they did could have led to significant learning on the part of all stakeholders: practitioners, researchers and policymakers.   But politics is a different kind of game. As it became more difficult to defend the Strategies publicly as a political success, so the government cut its losses. Pushed on to the back foot by the Select Committee report, the then Minister of State, Ruth Kelly, gave in to demands for an enquiry into the efficacy of the NLS approach to teaching reading in the early years and appointed as its Chair a civil servant who back in the 1980s had been part of a group within Ofsted advocating the use of phonics as the best remedy for poor performance in reading. Advice on the teaching of reading in the early years changed with the publication of the Rose report (Rose, 2006). The searchlights model that underpinned the National Literacy Strategy was replaced by the Simple Model of reading, with pride of place given to teaching synthetic phonics first and fast. Although the infrastructure to the NLS survived while New Labour remained in office, the

government itself increasingly put its weight behind other policies – the use of Academy schools to outsource system problems to alternative providers, with the risks for success or failure in turning schools around becoming their problem not the state's; and the re-framing of the education agenda around Every Child Matters, an attempt to bring other ways of addressing poverty and social disadvantage more directly into the policy mix, thus mobilising the resources of other agencies besides schools.

Over time the kind of clear target setting and close data monitoring associated with the form of top-down performance management that New Labour had championed in its early days dropped out of favour as a political strategy(Piatt et al, 2006). Faced with dilemmas that did not seem as if they could be satisfactorily resolved, the temptation grew to simply give the problem away. The Conservative party in opposition undoubtedly saw and fully understood the political risks that "governing by numbers" (Grek, 2009) brings with it. With their election to office, they drew a line under the New Labour approach. In his first speech to civil servants on becoming Prime Minister in 2010, David Cameron said:

> I'm not going to criticise everything the previous government did. Many of their intentions were right. Where they went wrong with reform was the techniques they used. Top-down. Centralising. Above all, bureaucratic. To improve public services, to get value for money, to deliver their stated aims, they set up a system of bureaucratic accountability. In this system of bureaucratic accountability almost everything is measured or judged against a set of targets and performance indicators, monitored and inspected centrally. The evidence shows this hasn't worked. … let me be very clear: I do not want you and your colleagues to think your role is to guarantee the outcomes we want to see in our public services - or to directly intervene in organisations to try and improve their performance …if you want to set targets, set new controls, impose new rules, don't bother because you're likely to get the red light. .. your job under this government is not to frustrate local people and local ideas, it is to enable them. (Cameron, 8 Jul 2010. https://www.gov.uk/government/speeches/pms-speech-at-civil-service-live)

But the consequence for teachers was that a whole way of working familiar from the Strategies would be dismantled around them with nothing very obvious put in its place. With the election of a Coalition government and the defeat of New Labour, the infrastructure of literacy consultants and the regular advice dispensed from the central Strategy team would soon cease. Within days the extensive collection of e-resources that teachers could consult, the product of almost 13 years of investment, were frozen, with a note appended saying "A new Government took office on 11 May 2010. As a result the content on this site may not reflect current Government policy". The range of resources designed to support system improvement would shortly be consigned to the National Archive at Kew. All that was left was a summary document setting out the full sweep of the Primary Strategies and their achievements, published under Crown copyright as a record of activity under the previous administration in 2011 (DfE, 2011). That was that. A National Curriculum built on political consensus and then re-articulated into a programme of curriculum delivery via the infrastructure of the NLS had been publicly disowned as a party political project.

A new look Department for Education website under the Coalition government made more prominent transcripts of Minister's speeches and press notices recording department news.   Policy pronouncements (funding for phonics resources, a reading progress check to be introduced in year 1) were mixed in with an eclectic and far smaller selection of resources on a much reduced range of topics, mainly pointing the reader away from the site to other organisations and providers that schools might want to consult. With the ending of the Strategies, the Department had stepped back from its role as the authoritative and reliable source of support for schools. Schools were on their

own now, acting as purchasers in a market for system improvement resources, choosing their own path, and following their own ideas, making their own fixes, and thus, of course, free to make their own mistakes too.

Why so much has changed, yet so much remains the same: autonomy and accountability in England
The current settlement in England can best be characterised in policy terms as high autonomy and high accountability. The government no longer directly supports schools by offering trustworthy guidance on what to do or how to do it under its own name. Schools are expected to act autonomously in making the best choices for themselves. Yet high stakes testing remains. If anything it has become more potent, placing more schools at risk of external intervention. This has become politically possible because by assuming no direct responsibility for the range or quality of support on offer, still less how schools act on that advice, governments are in a position to outsource responsibilities for school failure elsewhere. The increase in new providers in the market such as academy chains and free schools, coupled with the loss in many areas of a local planning structure, fully resourced and able to adequately support schools in a meaningful way, allows government to re-contract providers in the case of failure rather than take the blame themselves for the way things are.

The threat of schools being placed under new management if they do not meet benchmark standards is now enshrined in law. This seems to be the general direction of travel, giving the state's responsibilities away to anyone else prepared to take them on: if this organisation cannot deliver on the promise of high performance here, then another one will do. The narratives of disappointment the numbers create push more schools into the arms of different suppliers, without any evidence that they can really do any better by the pupils or the communities they serve. Democracy loses out as more schools slip out of local authority control. Meanwhile the threat of hostile takeover enshrines in the system a churn in staffing and instability in ethos and curriculum approach that militates against long-lasting improvements. The risks of disappointment that numbers on the public stage create are still there, but the responsibilities for fixing them now fall elsewhere.

At first sight the political risks of numbers-driven education reform may appear to be solved by combining high accountability with high autonomy. But a solution to one set of problems brings another set of problems in its wake. Performance management tools direct attention to the areas of most need on the assumption that more resources will then flow there. It is precisely managers' responsibilities to decide on what those resources should be. But by separating out, rather than holding together, the tools for assessing pupil performance and the responsibilities for curriculum resourcing and design, a whole new area set of tension points have opened up. Now that politicians in England no longer take any direct responsibility for producing the outcomes the system requires there is no real brake on the choices they make when deciding over the assessment instruments they still control. Institutionally and structurally, this is where things in England now stand.

This has led to a new and more paradoxical state of affairs. When first elected, the coalition government urged teachers to spend less time teaching to the test:

> we are concerned that especially in year six, there is excessive test preparation – with some children practising test questions for many weeks in advance of the tests. This is poor practice, and not even an effective way to do well in the tests, as compared to teaching well across a broad curriculum for the same period. We want to see whether there can be improvements to the current system. (*The Importance of Teaching - The Schools White Paper,* 2010, p48)

Accordingly, the planning and assessment structure of Levels that dominated teacher monitoring of pupil progression in the NLS was withdrawn. Yet at the same time, the potential freedoms that teachers were granted to imagine a curriculum detached from the planning strictures of the NLS became conditional upon any such curriculum matching the requirements of the new tests the government put in place.

Assessment has now become the main area of government reform. Yet the government no longer seeks to encode within the tests it has introduced a shared set of assumptions about how children learn or the best means of demonstrating the competences they have acquired (See Bearne, this volume). Lacking the responsibility to consider in detail how assessment, teaching and learning interact, and in the absence of the necessary tools to fully reflect on this, the gap between curriculum and assessment grows. Taken to its logical extreme, instead of building the curriculum and then deciding how it can best be assessed, the assessment tools themselves simply become the curriculum. This leads to a radical impoverishment of teaching and learning. In a high stakes testing environment, schools must teach children whatever the assessment asks of them. There seems no riposte to developmentally inappropriate assessment instruments that have not been standardised against actual children's performance and do not take children's age-specific developmental trajectories into account. The technical reports for both the phonics check and the grammar tests now in use demonstrate striking absences in these regards (STA, 2012; STA, 2013).

For policymakers in England, firmer control over the instruments of assessment and less interest in the detail of the curriculum seems to have created a new conversation within their domain. The belief that making tests harder will lead to better outcomes seems to have gained ground. Children who fail particular tests are now required to resit them. No attempt is made to adjust what is taught in the light of what children can actually do.

> Analysis has shown a difference in overall scores when comparing the term in which children were born, with younger children likely to do less well on the phonics screening check. This outcome was anticipated based on experience from other National Curriculum tests. However, the Government has high expectations for all children and therefore the policy does not take into account age for the purpose of setting standards on the phonics screening check. Therefore no attempt has been made to account for the age of the child.
> Standards and Testing Agency (2012) Year 1 Phonics Screening Check. Pilot 2011 - Technical Report. DfE. p9

Yet without using the tools psychometricians would employ to check on the reliability and validity of the tests put in place, their function in shaping meaningful teaching and learning in the classroom substantially diminishes. Meanwhile, in the absence of a clear story of what to latch on to and why, some schools will resort to their own systems of management by numbers, endlessly checking whether pupil performance measures up.

Conclusion: Re-imagining a purposeful and engaged literacy curriculum for all
Social imaginaries create possibilities for meaningful interaction in the world by reducing and stabilising inherent complexity whilst also leaving room to innovate and adapt (Jessop, 2009, p338). But in the English case the trajectory to numbers-driven reform has left schools struggling with the difficulties that numbers on the public stage produce, with less and less direct support from government on how to deal with them. This has done nothing to lift attainment. Meanwhile, the institutional uncertainty the punitive use of high stakes testing produces in schools only compounds what isn't working. Whole communities lose out under these conditions.

Re-imagining a purposeful and engaged literacy curriculum requires starting from a different place. The English case demonstrates that "governing by numbers" does not live up to the promises it makes. On the contrary, once standard solutions have been tried and found wanting, it sets in train a process of fragmentation and dispersion in which less well-attested "solutions" move through the system faster, and anything goes in terms of what happens next. Necessary protections for schools, for pupils and for the quality of teaching and learning, all weaken.

Many of the irreconcilable tension points in the current settlement stem from using statistical modelling to explore variation whilst expecting the outcomes to be better than average. This simply makes no sense. Variation in the system should not be a matter of blame. The resources schools get should not be conditional on the results they achieve. The architecture of high stakes testing needs to be dismantled. Other methods should be used to explain to parents how schools are extending pupils' capacities and capabilities through their teaching. Research evidence that can usefully inform practice should be shaped by the questions that concern teachers, pupils and communities the most. This requires re-gearing how knowledge is built, shared and applied within education. The clear purpose must be "to develop the most reliable and democratic ways of knowing, both in order to bridge the gap between ourselves and others, and to ensure that those who intervene in other people's lives do so with the most benefit and the least harm." (Oakley, 2000). That means finding new ways to take local conditions into account. There are precedents we can learn from. Scotland in particular has invested much more in trying to make such a model work.

To look across the history of literacy policy in England as I have done here is to highlight the fragility upon which social imaginaries often rest. A set of assumptions about what works in large scale education reform have been tested. The architecture of standardisation can do this much but not more. The real act of imagination is to decide where to go next.

References
Blanden, J., Greaves, E., Gregg, P., Macmillan, L. and Sibieta, L. (2015) *Understanding the improved performance of disadvantaged pupils in London*. London, CASE/LSE Working Paper 21

Coburn, C.E. (2006) Framing the Problem of Reading Instruction: Using Frame Analysis to Uncover the Microprocesses of Policy Implementation. *American Educational Research Journal*, 43.3, pp. 343–379

DfE (2011) *The National Strategies 1997–2011. A brief summary of the impact and effectiveness of the National Strategies*. London: Department for Education

Earl, L., Watson, N. and Katz, S. (2003) *Large-scale education reform: Life cycles and implications for sustainability.* London: CfBT.
https://www.researchgate.net/profile/Lorna_Earl2/publication/265758208_LARGE-SCALE_EDUCATION_REFORM_LIFE_CYCLES_AND_IMPLICATIONS_FOR_SUSTAINABILITY/links/5628fcff08ae22b1702edca5.pdf (accessed October 7, 2016).

Earl, L., Watson, N., Levin, B., Leithwood, K., Fullan, M. , Torrance, N. Jantzi, D., Mascall, B.

and Volante, L. (2002) Watching and learning 3. London: OISEUT/DfES.

Goldstein, H. (2001) Using Pupil Performance Data for Judging Schools and Teachers: scope and limitations. *British Educational Research Journal*, 27.4, pp. 433-442

Grek, S 2009, 'Governing by Numbers: The PISA 'Effect' in Europe. *Journal of Education Policy*, 24.1, pp. 23-37.

Hargreaves, A. (2002) Sustainability of educational change: The role of social geographies. *Journal of Educational Change* 3.3, pp 189–214. doi:10.1023/A:1021218711015

Harlen, W. (2014) *Assessment, Standards and Quality of Learning in Primary Education.* Cambridge Primary Review Trust Research Survey 1. http://cprtrust.org.uk/research/assessment/ (accessed October 7, 2016).

Harper (2015) Literacy teaching in the inner-London primary school: the shaping of professional practice, 1970-1979. Unpublished PhD thesis, Institute of Education, UCL

House of Commons Education and Skills Committee. (2005) *Teaching Children to Read. Eighth Report of Session 2004–05.* HC 121. London: The Stationery Office

Jessop, B (2010) Cultural political economy and critical policy studies. *Critical Policy Studies*, 3.3-4, pp. 336-356.

Lefstein, A. (2008) Changing Classroom Practice Through the English National Literacy Strategy: A Micro-Interactional Perspective. *American Educational Research Journal*, 45.3, pp. 701-737

Linn, R.L. (2000) Assessments and accountability. *Educational Researcher,* 29 pp. 4–16.

Lingard, B. (2010) Policy borrowing, policy learning: testing times in Australian schooling. *Critical Studies in Education*, 51.2, pp.129-147.

Lundahl, C. & Waldow, F. (2009) Standardisation and 'quick languages': the shape-shifting of standardised measurement of pupil achievement in Sweden and Germany. *Comparative Education*, 45.3, pp. 365-385.

Lupton, R. (2011) 'No change there then!' (?): the onward march of school markets and competition. *Journal of Educational Administration and History*, 43.4, pp 309-323,

Mitchell, J.C. (1983) Case and situation analysis. *Sociological Review,* 31.2, pp. 187–211.

Moss (2002) Literacy and Pedagogy in Flux: Constructing the object of study from a Bernsteinian perspective. *British Journal of Sociology of Education*. 23.4, pp. 549-558.

Moss, G. (2009) The politics of literacy in the context of large-scale education reform. *Research Papers in Education*, 24.2, pp.155-174.

Moss, G. (2014) Putting literacy attainment data in context: examining the past in search of the present. *Comparative Education*, 50.3, pp.357-373.

Piatt, W., Chau, S., Rajah, N. and Service, O. (2006) *The UK government's approach to public service reform – A discussion paper*. London: Cabinet Office

Rizvi, F. (2006) Imagination and the globalisation of educational policy research. *Globalisation, Societies and Education*, 4.2, pp.193-205.

Rose, J. 2006. Independent review of the teaching of early reading: Final report. London: DfES.

Sahlberg, P.(2007) Education policies for raising student learning: the Finnish approach. *Journal of Education Policy*, 22.2, pp. 147–171.

Sellar, S. (2015) Data infrastructure: a review of expanding accountability systems and large-scale assessments in education. *Discourse: Studies in the Cultural Politics of Education*, 36.5, pp 765-777. DOI: 10.1080/01596306.2014.931117

Soler, J., & Openshaw, R. (2006). *Literacy crises and reading policies: children still can't read!* London: Routledge.

Standards and Testing Agency (2012) *Year 1 Phonics Screening Check. Pilot 2011 - Technical Report*. DfE .

Standards and Testing Agency (2013) *English grammar, punctuation and spelling*. *2013 technical report.* DfE

---

[i] R000239406, Building a New Literacy Practice Through the Adoption of the National Literacy Strategy, 2002-3. ESRC
RES-000-23-0700, Remaking Literacy Policy for Schools, 2004-7. ESRC
RES-070-27-0053, Literacy attainment, data and discourse from the mid-19th century to the present day: a sociological account using mixed methods research, 2011-13. ESRC

[ii] For an example of the issues raised by taking one system recipe for reform into another, see the response to the introduction of a Common Core, an equivalent of the English National Curriculum, in the USA (Ravitch, 2016)