Putting literacy attainment data in context: examining the past in search of the present

Gemma Moss Professor of Education, Institute of Education, University of London

Abstract
This article examines the construction and design of literacy attainment data in the English school system in two different historical periods: the 1860s and the 1950s. These periods represent contrasting moments in the history of education in the UK when school structures and the governance of education varied, as did the design and interpretation of literacy attainment data, their relationship to pedagogy and the curriculum, and the administrative purposes for which they were deployed. By paying attention to the relationship between the form the data took and their mobilisation in administrative, professional and public discourse, this article will explore the role numerical data played in shaping education in the past and the tension points between data, discourse and social context that their use reveals. What can this tell us about the contemporary use of literacy attainment data and their current role in bringing policy and pedagogy into the same space?

Introduction
This article will consider two periods in the English school system when literacy attainment data played a prominent role in public discourse: the 1860s when the policy of "payment by results" was used as a key element in funding elementary education; and the 1950s when the 11-plus exam was used to determine entry to grammar schools. These two periods were chosen as they represent very different social circumstances under which knowledge about literacy teaching was shaped and enacted and different kinds of literacy attainment data were produced. The tools of statistical calculation available to make sense of the data also varied, as did the main questions upon which they were set to work. There are also continuities: in both the 1860s and the 1950s the literacy curriculum was conceived of as stratified and hierarchised in ways that could be represented numerically, even if the precise terms under which such hierarchies were constructed and validated were not the same. In approaching this diverse material I have treated the documentary sources from the past as evidence of different ways of conceptualising literacy attainment, and thinking about how it should be treated, pedagogically, administratively and in public discourse. Rather than adjudicate between the various ways in which the data are constructed and mobilised, I've sought to understand why and how particular perspectives made sense to the community that voiced them at the time, and the conditions that made them vulnerable to review, dislocation and displacement.

This approach has some resonance with work on epistemic communities interpreted as "a network of professionals with recognized expertise and competence in a particular domain" working with a "shared belief or faith in the verity and the applicability of particular forms of knowledge or specific truths." (Haas 1992, 3). Haas argues that the knowledge or information such communities amass has a heuristic quality, tied to the particular contexts in which such knowledge is generated:

> "It consists of depictions of social or physical processes, their interrelation with other processes, and the likely consequences of actions that require application of considerable scientific or technical expertise. The information is thus neither guesses nor "raw" data; it is the product of human interpretations of social and physical phenomena." (Haas 1992, 4).

Haas suggests that such information gains traction in policy circles under conditions of uncertainty when established ways of doing things have broken down, yet the rapid relocation of such knowledge into new contexts of use may lead to unpredictable and unpredicted outcomes (See also Davis Cross 2013). Who is entitled to make what claims about the kinds of knowledge they have to hand is a key feature of each of the case studies from the past. At the same time, due recognition

has been given in the analysis to the different conditions under which such specialised communities of knowers might form (Maton 2013) and the different form boundaries between institutionalised forms of knowledge that may have held at the time. The distinction Schriewer (2014) makes when characterising the 'socio-logic' of reform-orientated studies designed with an immediate practical purpose in mind by "contrasting modes of perceiving and thinking, shaped in accordance with different frames of reference – with scientific search for truth or with a sub-system's needs for self-description and guidance" (92), is also important here. The emergence and use of particular forms of literacy attainment data in the past has been analysed in this light.

Interpreting numerical data from a qualitative point of view
The data considered here stem from a larger research project, funded by the ESRC, which explored the interplay between public discourse and statistical descriptions of the social world in three different periods, using literacy attainment data as its primary focus. The project as a whole focused on the role of three administrative bodies: the Committee of Council on Education (CCE), the precursor to the Department of Education, which administered grants to elementary schools in the 1860s and collected attainment data as a means of paying schools; the National Foundation of Educational Research (NFER), a research body which was initially based at the Institute of Education, University of London in the 1940s, and began to research as well as design standardised tests that could be used as part of the 11-plus; and the OECD, the international body located in Paris whose design, collection and use of PISA data has become a powerful means of shaping policy advice to governments in the contemporary period.   Only the first two cases will be considered here.

I came to this project, not as a historian or a comparative educationist, but as an ethnographer of literacy pedagogy and literacy policy, with a keen interest in how these two different spheres of practice interact (Moss, 2009). In exploring the use of literacy attainment data in the past I was not expecting the past to explain the present, in terms of an increasingly secure knowledge base building seamlessly and inevitably on what had gone before. I was interested in the different logics at play which rendered the data plausible and intelligible in each period, within particular frames of reference, whilst bringing it under scrutiny and disputation elsewhere. Although secondary sources were useful guides to institutional structures and key players (Bishop 1971; Sutherland 1984; Wooldridge 1994; Stigler 1999) I did not rely on the narrative they constructed to see the key terms of the debate, but rather sought out and reassembled from primary sources the points of agreement and dispute in contemporary discourse over the nature of the numerical data collected and their relationship to the curriculum. What did the data represent for contemporary audiences as viewed through the administrative and or pedagogical purposes that they served at the time? The past emerges from such an enquiry as composed of multiple narratives, sometimes directly competing for attention, sometimes borrowing from each other, sometimes interlocking and overlapping, but essentially developing in distinct spheres of practice in which the necessary knowledge base to shape or interpret the data emerges and develops over time under that setting's specific conditions of use (Cowen 2002).

Most research effort was spent on the 19th century case, which also involved the longest sweep of time. To understand the social organisation of the literacy curriculum in 1860s elementary schools in Britain involved tracking back from the pedagogical resources for teaching literacy then in use to their origins in the mid 18th century, as well as consulting contemporary advice to school teachers on reading pedagogy, offered in professional journals or manuals and exemplified in exam syllabi for teachers. I also explored the origins and development of the sponsoring organisations that funded and managed elementary schools, using administrative sources. To place the policy of payment by results in context involved reading the CCE's annual reports from the 1850s onwards, considering government enquiries on the funding and functioning of elementary education, and reviewing parliamentary speeches and the public petitioning and commentary that arose after the policy was

announced in 1861. The traces of public debate over whether and how to fund education for the poor were also followed back to the beginning of the 19th century[1]. The administrative data the CCE collected on attainment and other aspects of the school system were also tracked over time, with attainment data on reading, writing and arithmetic from 1864-1872 entered into SPSS and analysed. The 1950s case was more contained. I tracked the research the NFER undertook into secondary school selection and the use of standardised tests, using the archive of materials available in the Institute of Education, University of London. I also consulted Brian Simon's archive on Intelligence Testing held at the same place. This contains a good deal of contemporary press coverage and other source materials that instantiate the public debate. In addition I reviewed contemporary publications on selection and examination whilst tracking back the development of standardised testing of the kind used in the 11 plus from the NFER via the work of Godfrey Thomson and Philip Vernon to Cyril Burt and the Binet-Simon tests.

Numbers are seldom as transparent as their proponents suggest (Waldow 2001). By studying two cases of the use of literacy attainment data at different points in the past this article hopes to re-steer a conversation about what such data can or cannot be taken for in the present. How do literacy attainment data take on different resonances as they travel across different communities of practice and into different social and institutional contexts? What are the interactions between contexts that shape the trajectory to the data's use? What are the circumstances that test the data's function or lead to their demise? These questions will be considered through two narrative case studies of the formation, interpretation and uses of literacy attainment, bringing out the points of agreement and dispute in contemporary discourse about the data in the 1860s and 1950s.

The policy of payment by results in 19th century England: numerical answers to dilemmas in policy and pedagogy

System-wide attainment data were first systematically collected in England in the 1860s as part of the policy known as "payment by results". Prior to this, the Committee of Council on Education (CCE), the body charged with oversight of government expenditure on education, had been using a system of matched grants to help charitable sponsors with the costs of building elementary schools to educate children from the labouring poor. The grants supplemented monies raised from local subscriptions and school fees by the sponsors, religious groups organised on broadly sectarian lines. Later the government began to offer additional monies for system improvements: to support the purchase of equipment; to augment the salaries of staff qualifying as Certificated Teachers by passing professional examinations; and for the employment of pupil teachers, using a fully government- funded five year apprenticeship scheme. The CCE kept track of their expenditure in annual accounts, recording how much money had been spent in total, and what it had bought. The main expenditure was broken down into numbers of schools, with floor space calculated as square footage per pupil; numbers of teachers, including what proportion were Certificated; numbers of pupils in average attendance, by age; and numbers of pupil teachers, all set out by sponsoring organisation (See e.g. The Report of the Committee of Council on Education, 1862-63).

These funding arrangements had grown up in an ad hoc manner at a time when parliament continued to vote down successive proposals to fully fund a national system of elementary education, reflecting the deep ambivalence felt at the time over whether it was the state's business to intervene directly in the education market on behalf of those too poor to purchase education for themselves, and with little appetite to shoulder the costs from taxation (Green, 1990). Meanwhile, as the sponsoring organisations continued to expand their provision, often competing against each other in ways that made efficient planning of school places difficult, government expenditure on

---

[1] Details of the main primary sources consulted and their range can be found at
< http://libguides.ioe.ac.uk/literacyattainment>

education continued to rise (BPP, 1865). With no clear idea of the scale of the need matched to local population density, or when the increase in demand might slow, the government became increasingly anxious to understand what precisely what they were getting back for the monies spent. The Newcastle Commission, set up in 1858 to "Inquire into the Present State of Popular Education in England, and to consider and report what Measures, if any, are required for the Extension of sound and cheap Elementary Instruction to all Classes of the People" decided, broadly speaking, that the answer was not enough. In particular they reported too few children leaving school with a lasting competence in reading, writing and "cyphering" (arithmetic).

> "If a child of 10 years old, who has attended school with moderate regularity for four or five years, can hardly read and write and cannot cypher to any useful purpose, it is very hard to call upon the parent to keep him at school four years longer and to tax him with gross selfishness and ingratitude because he does not choose to forego a large addition to the family income in order to do so. ... If a child is ignorant of these after four years schooling, his parents may well be excused for thinking the experiment has lasted long enough."
> Newcastle Commission, BPP, 1861, p177

The quote above hovers between two competing explanations for this state of affairs: that children were not attending schools long enough to learn, or that the teaching in schools was not sufficiently efficacious to keep them there. The latter charge was based on the suggestion that in classrooms organised on monitorial lines, poorly trained pupil teachers carried too much responsibility for teaching the early stages of reading, whilst the teacher spent most of their time with the ablest pupils, neglecting the basic levels of instruction to teach the highest class a more rewarding range of subjects (Ibid, p 247).

In response to the Newcastle report and the concerns it raised, in 1861 Robert Lowe MP, then Vice President of the Committee of Council on Education, brought put forward proposals for a new means of funding education using a system of payment by results (Sylvester, 1974, pp 40-57; Hansard, 1861-1862). This was intended to place the financial settlement for education onto a sounder basis, ensuring more tangible outcomes from the money spent. At the time and in much later commentary Lowe was accused of opting for a system that would destroy the quality of educational provision (Kay Shuttleworth, 1861). In fact there is little evidence that he opposed education for the working poor, rather more to suggest that he objected to the lack of focus on ensuring that all pupils would leave elementary education literate and numerate, and was suspicious of the sectarian motives he thought guided too much of the activity of the schools' sponsors (Sylvester, 1974; BPP 1865). His solution was to replace the complex system of individual improvement grants with a single payment to schools made according to the number of pupils who passed examinations in reading, writing and arithmetic with school managers able to spend the money as they wished.

The examinations were to be conducted by school inspectors during their annual visits. Administrative efficiency would be served if the tests could be completed in the single day the inspector was there, and different calculations of the numbers of individual pupils an examiner could examine under the most propitious conditions were used to help determine the precise form they should take. There were few attempts to standardise the criteria against which Inspectors were expected to make their judgements (CCE, 1862-63, pp.xxi-xxii). Instead the sequence of difficulty expressed in the textbooks commonly in use to teach reading were now used to define the successive levels in the examinations (Michael, 1987; Tilleard, 1860). (Lowe was certainly familiar with the most popular series, the Irish Reading books, originally published in the 1830s (Lowe, 1848). But Lowe's examination sequence also redefined the end point of the curriculum as the capacity to read secular texts of a kind that circulated outside the school and would provide reading material

well into adulthood. The Revised Code as passed through parliament stipulated that for the purposes of examination in reading:

> Children should read aloud:
> Standard I   Narrative in monosyllables
> Standard II   One of the narratives next in order after monosyllables in an elementary reading book used in the school
> Standard III   A short paragraph from an elementary reading book used in the school
> Standard IV   A short paragraph from a more advanced reading book used in the school
> Standard V   A few lines of poetry from a reading book used in the first class [the highest class]in the school
> Standard VI   A short ordinary paragraph in a newspaper, or other modern narrative
> (BPP, 1862)

This broke with the trajectory in use in most schools earlier in the century, set out in Joseph Lancaster's monitorial system, which stratified the literacy curriculum by word length, and expected the highest class to make the transition to reading the bible:

> Class I.   The Alphabet
> Class II.   words or syllables of two letters
> Class III.   ditto,   three letters
> Class IV.   ditto,   four letters
> Class V.   ditto,   five letters &c.
> Class VI.   Reading or spelling lessons of two syllables and Testament
> (Lancaster, 1821)

Lowe's intentions in terms of the level of proficiency he wished children to acquire were clear, but it was much less certain whether these goals could be achieved in practice. There were various impediments to which he had paid little heed. Teachers and school managers were keenly aware of the irregularity of attendance which dogged schools at a time when education was not compulsory, when children and families might move frequently in search of work and when poor health was rife (M.U.R, 1861). Professional knowledge of how to teach reading was also under-developed, with many classrooms still using the monitorial practice of asking pupils to read out loud and then spell letter by letter each individual word in the passage they were studying, a practice that encouraged rote-learning as a key pedagogic strategy (Commissioners of National Education in Ireland, 1835). By contrast, Lowe's stipulations altered expectations about the speed at which students would muster sufficient skill to deal with passages of increasing length and complexity.   New textbooks followed that sought to match the sequence in the examinations more precisely, supplying more reading material at each of the required levels with advocates of Phonics and Look and Say arguing the merits of their respective approaches   (Laurie, 1863; Vincent, 1997). But the advent of new textbooks did not in itself dislodge older routines and practices.   From the point of view of many in the teaching profession the examination sequence Lowe stipulated looked unachievable. Anticipating that too few pupils would be able to meet the reading demands in the exams, many predicted a catastrophic collapse in the element of government funding that supported their activities (Kay Shuttleworth, 1861).

On his side, Lowe calculated that 75% of children would be able to pass the examinations for which they were submitted and, based on this, fixed the fee per pupil at a level that he thought would maintain current school grant income. In the event, the pupil pass rate in each subject and standard of examination met or substantially exceeded this expectation (in reading it was consistently above 90%), but school incomes dropped as far fewer pupils than Lowe anticipated had attended long

enough to be entered for the exams, a stipulation designed to ensure that schools were paid for the work they had actually done. The reduction in income, the pressure to maximise the number of children submitted for examination, the changes in the conditions of teachers' employment that accompanied the new dispersement of funds, all alienated the profession. This created lasting difficulties that would rumble on for much of the rest of the century.   Stories about the vulnerability of teachers to the capricious judgements made by inspectors, who would surprise a class by examining them on an unseen text rather than the passages they had been practising from a familiar reader, exemplified the sense of injustice (M.U.R, 1861). Meanwhile, government monitoring of the numbers of pupils of different ages submitted for each standard of examination as against the numbers of pupils in average attendance and in class on the day of the inspector's visit began to make clear how irregular attendance really was. Despite initial attempts to suggest that teachers were gaming the system by deliberately submitting older and more competent students for the easiest of the exams, the net result was a recognition of the difficulties faced and a change to the funding formula. More money went to schools in recognition of average attendance not exam passes, and when school boards were created with control over education in their area in the 1870s, greater encouragement was given to introducing byelaws making school attendance compulsory.

In public, as opposed to professional discourse, attitudes towards the policy changed over time. When Lowe introduced the policy of payment by results into English schools there was a general enthusiasm for examinations as instruments of reform, on grounds that they could be used to reward merit, encourage the competitive spirit amongst examinees, and attract public notice as a spur to effort (Layton, 1973). Examinations in the civil service and in the army were considered the best means of moderating the exercise of patronage and curtailing nepotism; in medicine and in teacher training they had been introduced to develop and uphold clear standards of professional knowledge; in the Arts they had been intended to stimulate talent through the award of prizes, encouraging the numbers taking up new skills. But a counter discourse treated their effects with suspicion. Many feared examinations encouraged cramming and mechanical learning rather than identifying real accomplishment, producing the simulacrum of ability in return for a minimum of intellectual effort; they narrowed the curriculum to the contents of the textbook, and minimised the benefits of the extensive contact between teacher and pupils that stimulated real thought. Ahead of the introduction of "payment by results", Joshua Fitch, then principal of Borough Road Normal College, later an HMI, warned against the "examination mania" of the time in an address to the National Association for the Promotion of Social Science:

> we find ourselves indirectly encouraging cramming, diminishing independence and originality of thought, substituting routine for invention, memory for intelligence, and the dulness of the mere student for the freshness, life, and animation of the energetic teacher Fitch, (1859) *Examination Schemes, and their Incidental Effects on Public Education.*

 As the century wore on school examinations became increasingly associated with excessive demands made by bureaucrats on teachers and on pupils, giving little of real value in return. The "over-pressure" panic of the mid 1880s identified injurious impacts on health from the examination process on both teachers and pupils. Nervous collapse and brain-swelling from too much mental exertion were linked in the medical literature and in the press to the pressure of overly-taxing examinations and popular demonstrations against elementary education were mounted on this basis (Galbraith, 1997; Middleton, 2004). If payment by results had solved an immediate dilemma in funding in the 1860s, easing the way for the state to take more control over the curriculum and ultimately of schools, any perceived advantages had long since receded into the past when in the 1890s the policy was finally abandoned and the state took over fully funding primary education for all.

The policy of "payment by results" had paradoxical consequences. It gave the government much clearer oversight of expenditure on education. Control over the content and conduct of education moved from the sponsoring organisations to the state. But funding education through examinations did not alter pedagogy. Rote-learning continued to exert its hold. Moreover, although the turbulence in attendance became easier to see by tracking how many children were submitted at which age for which exam, the economics of child labour continued to keep children out of school, setting up a battle over compulsory attendance between working class communities and the state (Middleton, 2004). The policy resolved the relationship between the state and the sponsoring organisations that had begun opening elementary schools, but it was not the main lever in bringing more children into school.

Testing and the tools of science: Rationing access to grammar school places in the 1950s
By the 1950s attainment data had come to occupy a rather different place in the English education system. The immediate policy problem post-war was not how to fund schools but how to extend the time children spent in education and what form that education should take. Since the early 1900s control over education planning had largely fallen to the Local Education Authorities (LEAs) (Sutherland, 1984). Their different strategies for extending the period children spent in school had led to a mixed system of secondary education, in which grammar schools continued to operate as relatively independent institutions alongside "higher grade" elementary schools, that provided an extended education for pupils who wished to stay on beyond the age at which they could leave for work, in turn co-existing with other more dedicated forms of post-primary education to which elementary pupils could transfer in preparation for "the various branches of commerce, industry and agriculture" (Morrish, 1970, p53; see also Dempster, 1954, p5). Within this diversity of supply, the grammar school took pride of place through its academic curriculum conceived of as preparation for university and the professions, and its association with middle-class students who expected to follow that route. When the government decided to raise the school leaving age to 15 in 1944, it advocated pupils transfer at 11 years of age into a new tripartite system of secondary education, variously catering for the "age, ability and aptitude" of pupils. In effect this meant maintaining the existing selective grammar schools whilst consolidating other provision into technical or modern schools. All of this intensified the scrutiny of the selection processes by which children gained entry to grammar schools, an already contentious issue, made more so as pressure on places increased (Sutherland, 1984).

In the 1950s the examination used to determine entry to the grammar schools, generally known as the 11-plus, was neither centrally organised nor benchmarked, being instead locally set and administered. Its component parts varied according to the weight individual grammar schools or local authorities placed on written examinations, school-based interviews, standardised test instruments and teacher recommendations, either in combination or singly (Sutherland, 1984). Operating against a background of considerable variation in the number of grammar school places available, whether for boys or girls, both regionally (there were places for only 8% of pupils in grammar schools in Gateshead, but 60% in Merioneth), and within particular districts (there were places for 40% of pupils in one part of the West Riding of Yorkshire, only 10% in another) (quoted in Wooldridge, 1994, p261. See also Yates and Pidgeon, 1957, p23), there was no fixed pass mark in the 11-plus. As Wooldridge commented:

> "In general most educational authorities agreed that the minimum standard of ability and attainment required for admission to grammar schools was an arbitrary standard determined by the accommodation available. The regional distribution of places bore no relation to the regional distribution of abilities or occupations. (Wooldridge, 1994, p261)

The 11-plus took place, therefore, against a background of locally-rationed supply in which only some were expected to succeed. Accordingly the main priority was to use the examination process to establish a rank order amongst the local candidates so that they could be admitted in a descending sequence of merit until such time as places were filled and the local cut-off point was reached. The main emphasis fell on being able to accurately pick out those who might benefit most:

> "the purpose of the examination is the selection at the age of 11-plus of children fit to profit by secondary education. The importance of accurate selection is vital and the main business is to get the right children." Supplement (1938) to Board of Education, Schools and Free Places. Quoted in Wiseman, 1961.

Accurate selection proved surprisingly difficult, both from the point of view of the schools and also the potential applicants (Yates and Pidgeon, 1957, p12). In the early stages, children applying to grammar schools from elementary schools were effectively being asked to transfer from one form of curriculum to another with few concessions made in the entrance exams to the difference in subjects studied at what depth. (Gordon and Lawton (1978, p200) reproduce a copy of a rejection letter sent to an elementary school pupil in 1901 who had failed on the Latin test as well as showing insufficient knowledge of Geography and Arithmetic. His better performance in History was not judged to compensate for this.) Trying to disentangle ability from attainment based on prior teaching remained difficult. Attempts to adapt or change the nature of the exam so that it was less tied to a specific curricular content did not resolve these issues, when success in the grammar school depended upon doing well with the precise curriculum it offered (Bryce Report, Vol iv. R P Scott evidence, Qu 13,990; 14,003-041, quoted in Montgomery, 1965). Candidates also needed help to prepare for the exam if they were to do themselves justice, but schools varied in their ability to offer that support. Moreover, if pupils could be substantially helped to improve their performance, this in turn raised the question of when help shaded into coaching, and whether coaching might then obscure which candidates had real talent, thus inhibiting grammar schools from selecting the "brightest and best" after all (Wiseman 1961). Some schools in remote rural districts with no history of successfully submitting pupils for the exam simply put no candidates forward, a situation that the educationist, Godfrey Thomson, deplored (Thomson, 1921).

Questions over the reliability of the selection procedures seemed borne out by the attrition rates: some pupils awarded places left before completing the course; others demonstrated only mediocre performance, observations that underpinned one of the first programmes of research undertaken by the NFER after its foundation. The potential waste of talent from "inefficient" admission procedures was exacerbated by the social prestige associated with grammar schools that saw middle class parents making as much effort as they could to get their children in. This mattered less when middle-class students entered via a separate fee-paying route, but once all grammar school places became state-funded after 1944, with the exam the main gate-keeper, the potential for middle class parents to exploit opportunities for coaching or use personal networking to gain recommendations for entry intensified. Meanwhile, financial pressures continued to prevent many students from working class families from taking up the places they had been offered or, once there, inhibited them from staying on rather than leaving at the first opportunity. (The London education authority, the London County Council, took this into account by formally eliciting and recording parents' wishes for secondary education on the form they used to collate the relevant information for the 11-plus.)

The new forms of intelligence tests that began to be developed in the 1920s seemed to promise to resolve these difficulties through the deployment of different testing techniques. Intelligence tests represented a new departure in the conduct of examinations through the application of the latest statistical concepts and methods (the normal curve, standard deviation, the method of least squares, correlation coefficients and regression to the mean (Stigler with Kruskal, 1999)) to the task

of identifying distinctions in attainment and abilities (Burt, 1962; Thomson, 1924; Vernon, 1960). Thomson considered the main advantages of these statistical techniques rested with their capacity to identify both wide range in attainment and continuity over the range in any particular dataset:

> There is always a danger that the use of class names, such as defective, normal, clever, genius etc may lead to the belief that these classes are quite distinct. Thus children were once rather looked upon as falling into a large group of normal children, all very much alike, and two smaller groups, of defectives at one end and, at the other, of geniuses, clearly marked off from the normal. Now this is not so, as experiment and statistical observation have clearly shown. If men or children are ranked for intelligence they form as unbroken a series of steps from one end of the line to another as they do in the case of height. There is not a class of defectives at one end and of geniuses at the other at all clearly marked off and separated sharply from the others. There is continuity, there are no gaps which cannot be filled in, if we increase our numbers. (Thomson, 1924, 171)

Taken up in the context of the 11-plus, advocates suggested a more accurate order of merit could be established through the application of the new statistical techniques, thus providing a better basis upon which places should be awarded at grammar schools (Dempster, 1954). Yet the specific techniques of intelligence testing failed to replace existing approaches to selection in the 11-plus, being more often combined with them. Gillian Sutherland's (1984) painstaking and detailed research into the development and use of intelligence tests as part of LEA arrangements for the 11-plus from the 1920s to the 1940s shows how little headway they made against other methods during this time, even though they became the central rallying point in campaigns against the 11 plus mounted in the 1950s (Simon, 1953; Banks, 1953).

At first sight this seems odd. The tests' scientific credentials were based on statistical procedures applied to data obtained through "actual experiment" (Thomson, 1924 p181), with children of different ages tackling the same tasks or questions. Examining the distribution of marks and their dispersion around the mean using statistical techniques made it possible to draw comparisons between individual children with apparent precision (Dempster, 1954; Thomson). Yet in practice these calculations were not straightforward. Not all of those devising tests applied the rigorous standards to test development pursued by Thomson's institution, Moray House, and the NFER, which entered the test market shortly after its foundation (Wooldridge, 1994). To enable the calculations to proceed, questions needed to have clear right /wrong answers. This narrowed the focus of what could be tested whilst leaving open the relationship between test performance and the requirements of the curriculum. Practically speaking, the statistical tools provided little advice on what to do if the intelligence test results conflicted with other evidence, such as the candidate's capacity to write an extended essay, even though this remained so central a feature in grammar school education. Moreover it was also clear that children performed differently on different tests, and even the same tests sat at different times – there was not a constant rank order that carried over from one context to another. Whilst statistically speaking some of these difficulties could be addressed by aggregating data from a number of tasks using techniques such as factor analysis, it was a matter of dispute between statisticians as to what the outcomes really signified, and in particular whether differences in performance could be harmonised to generate a general measure of intelligence, or represented different kinds of intelligence that required separate recognition (Wooldridge, 1994, p 95, 291).

The statistical tools did not really resolve the social arguments either. Thomson, for instance, who did most to produce reliable test instruments for school selection purposes, hoped to improve the opportunities for bright students from poor backgrounds to find their way into the grammar schools, a route he himself had followed. He recognised the importance of whole group testing as the most

reliable means of benchmarking average attainment within a particular region, against which the judgements involved in selection for the available grammar school places could be made. In practice this suggestion was often resisted by teachers and LEAs as inappropriate given only a few pupils would be expected to transfer to grammar school. He hoped that accurate measurements would give everyone the same chance to be identified for selective entry regardless of their background. Burt, by contrast, having done much to set the original direction of travel in "intelligence testing" with his work on *Mental and Scholastic Tests*, became increasingly interested in genetic inheritance, and in proving via intelligence tests the role he thought it played in determining fixed abilities (ibid pp92-5).   The different kinds of arguments that could be mounted via the statistical data were thus refracted back through and used to re-articulate contrasting positions in public discourse. Rather than moving the whole process onto neutral ground, the contradictions in what the test data showed were increasingly mobilised and deployed by those disputing how the task of selection at 11 was carried out. An article on the subject in the Daily Herald began:

> Do you know how many beans make five? Good then I must tell you that a great many serious minded people are getting rather heated about that question.   ... For precisely that sort of stupid question is playing a large part in deciding whether 11 year old children go to grammar or secondary schools. (Daily Herald, 4.12.53)

The article continued with the sub-headings: Scrap the tests; Answer these; Brains – or wits? Teachers say.

The more time and resources the educational community committed to researching these issues, the less certain it seemed that the available tools could answer the immediate questions in selection. After extensive research following up pupils who had passed the eleven plus and those who had not, the National Foundation for Educational Research (NFER) could only hazard that intelligence tests reliably identified the top 10-20% of children, and that beyond that one entered a "border-zone" where statistically speaking one could not discriminate between pupils attaining a relatively broad range of marks (Yates and Pidgeon, 1957, pp144-5). Other factors including a willingness to work hard had to be taken into account in identifying who would flourish in grammar schools, whilst the damage done to those with similar marks who had failed the exam and were sent to secondary moderns troubled many. As the psychologist, Alice Heim, commented in a letter to the journal *Education*,

> "It is the exaggerated claims of the test devisers and theorists that cause the trouble. Who but a psychologist would discover that "personality and interests" influence test results? ... To refer to this fact as "a factor" and seriously to designate it "the X factor" is to point the absurdity of trying to reduce school children .. to an enumeration of measurable "factors" and sardine-tin them into a theory." Heim, The Psychological Laboratory, Downing Place, Cambridge (Heim, 1957)

 As the tide turned in favour of greater social equality, instead of planning to expand grammar school places or refine the test instruments, some LEAs preferred to opt out of testing and proposed more comprehensive schools instead (The Times, 20.11. 52.; News Chronicle, 10.1.53)

In the 1950s, numerical data come to the fore via the advocacy of a specialist community interested in the application of statistical techniques to educational processes. The data were designed to discriminate securely between the capacities of individual pupils without reference to the curriculum or to processes of teaching. Yet it was never entirely clear whether this claim was really true or how far it helped schools decide who would benefit most from their curriculum. Over time the tests' administrative application had changed from identifying children who would not benefit from

ordinary schooling to selecting those most likely to do well in grammar schools. But the techniques could not resolve the issues that had already brought the processes of selection for grammar school into public dispute: the geographically uneven distribution of places and the shortage of supply in relation to demand. Instead the uncertain relationship between the results from the tests and the judgements of teachers provided ammunition for those arguing that selection at 11 was neither efficient nor necessary, a campaign fuelled by a new egalitarianism that would end with the introduction of comprehensive schools.

From past to present: continuities and differences in the uses of numerical data
Neither of the cases from the past provide an exact template for the uses of literacy attainment data in the present. In both cases the form the data took provided partial solutions to immediate administrative difficulties: in funding and in planning for growth in the 1860s; and in organising entry into secondary education in the 1950s. In both the 1860s and the 1950s the data also became the subject of very visible and vociferous campaigns, acting as rallying points for teachers and the wider community in very public arguments over the use of the data and the administrative arrangements they addressed.   These conditions are not directly replicated in the present. The attainment data that now play so large a part in the organisation of education have not become the object of such vigorous campaigning, just as it is less immediately obvious what were the administrative difficulties or structural change they were initially intended to address. Yet the problems they now define loom large (Lefstein, 2013). Indeed, one of the oddities of the current context is the extent to which the data now come first: they have become the primary signs of system health which drive other administrative and pedagogic activity in reaction to whatever the data seem to show.

Placing system data first has become a general policy prescription, strongly advocated by the OECD, arguing that this brings better system planning and improved system quality (OECD 2011; 2013). Such an approach assumes, of course, that the relationships between attainment, curricular progression and quality of teaching and learning can be adequately captured in numerical form. System-wide comparison of the data will then provide the most appropriate stimulus for professional reflection, whether on the part of administrators, policymakers, or teachers, with the data acting as an appropriate catalyst for change. Whilst there has been some trenchant criticism of these propositions within the academic community (Sahlberg, 2011; Biesta, 2009; Luke, 2010; Lawn and Ozga, 2009), public discussion and debate have been much more muted. (See Landahl and Lundahl, 2013).

The historical case studies explored how the data interacted with diverse forms of knowledge already circulating within different communities of practice. This helped set the parameters to how the data would be interpreted in different contexts of use. Separating the form the data take from the discourses in which they are embedded is a useful principle here.   In the current context, two particular concepts from statistics are used to structure analysis of attainment data: modelling distributions in the data to identify variations in outcomes; and modelling progression in the data, using linear regressions to standardise and then compare rates of growth.   These statistical techniques are not necessary components to teachers' work in school. Far simpler means are available in real time to distinguish between children's current capabilities as they tackle tasks in class, helping teachers foster pupil learning through immediate feedback and plans for future teaching (Sahlberg, 2009). But they do create strong points of comparison at system level which in politicians' hands become powerful ciphers of education quality. In the process the data are transformed: ways of summarising and interpreting data statistically turn into matters of hard judgement. Natural variation in the distribution of scores within a population – data that in the 1950s would have been used to determine who went where within a selective system of secondary education - now create criteria by which to conclude that schools and pupils in the lower range could and should do better. Outliers in a statistical dataset that by definition occupy a position in the tails

of the distribution are treated as signs of unacceptable failure or alternately promoted as achievable models of success. In the latter case the assumption is that any school could do as well, if only they tried harder. The average, a descriptive term for the mid-point in a dataset that can be variously calculated as mean, median or mode, becomes the politician's "unsatisfactory". Clustering around the mean, which statistical analysis would predict, is treated as evidence of under-performance. Interpreting the statistical data in this way produces a powerful discourse of disappointment. Approached in this light, progression modelling is used to create a sense of failure whenever pupils fall beneath a line that must place 50% of the population above or below its midway point. Under these conditions and through politician's use, statistical tools designed to explore variation in any dataset now serve to highlight the distance in reality from the uniform outcomes politicians crave. Precisely because the statistical modelling works with relative rather than absolute values, the data establish ever moving targets for schools to chase, setting up system goals that by definition cannot be met.

Accountability cultures in English schools that currently drive the generation and use of test data are muddling up what the statistical analysis consistently shows - that attainment varies – with expectations that under the right conditions all could achieve at the level of the best. In the English context the increasing politicisation of curriculum and assessment as tools for system management are making it easier for key decision-makers to exploit the potential for the data to create failure rather than spread success. This has become part of a political endgame which may yet push more schools into the control of private sponsors who are allowed to play by different rules: schools that do not meet benchmarked levels of achievement are subject to inspection and "special measures" that can lead to imposed academy status, reallocating the school from local authority control to the control of private sponsors. This process happens at central direction with little regard for eliciting local consent from parents or the wider community. The numbers on their own are treated as definitive information that justify the move. They do not become talking points between a range of parties entitled to interrogate precisely what they mean and query whether they are really sufficient for the purposes for which they are used (Goldstein and Leckie, 2008; Leckie and Goldstein, 2009).

There are serious problems for the quality of education when assessment data are deployed in this way. The more assessment data are harnessed to audit outcomes, the more the space for the exercise of teacher judgement in profitably interweaving the processes of learning and teaching in the classroom diminishes (Lefstein and Snell, 2014). Teaching reverts to faithfully delivering whatever is next on the curriculum list whilst successful learning is re-defined as producing the data that meet whatever benchmark is measured by the test. As the 2010 Schools White paper on *The Importance of Teaching* said

> 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, so that parents have the information they want and schools can be properly accountable for pupil progress, without encouraging over-rehearsal of test questions. (DfE, 2010)

So far the new forms of testing the present administration in England has introduced do not suggest it is living up to its word or trying to break out of the discourse of disappointment that current uses of system-wide statistical data have helped create.

Conclusion
The historical cases suggest that the story attainment data tell depends upon the form they take *and* the contexts in which they are interpreted and understood. Attainment data do not stand apart

from the public and political fray into which they are introduced, acting as neutral arbitrators between opposing points of view. They gain their prominence through multiple interactions between different knowledge communities that unfold in time and space. In comparison to the past it is striking that there has been so little public scrutiny and opposition to the use of attainment data in the present. Part of the reason may simply lie elsewhere in the broader social and economic landscape. The use of statistical data to produce a discourse of disappointment is happening against a larger background where the relationship between the credentials students carry away from education and the functioning of the economy have become more uncertain. In England more of the costs of education are being passed to the individual from the state. In an hour-glass economy characterised by high levels of income inequality and job insecurity, it is now more doubtful what the qualifications the education system provides will actually buy. Under these conditions, the proposition that education is not working, so effectively demonstrated in politicians' use of statistical data on education performance, may reverberate with rather than cut against, perceptions that resonate within the wider public sphere.

The historical data shows that attainment data do not operate as clear-cut multi-purpose administrative tools that can unilaterally fix pedagogy, solve funding dilemmas, sort students and deliver the right resources to the right place. Their construction is more context-specific, more time-limited and contingent than that. Looking across each of the three cases, the limitations of the data and what they can encode in terms of attainment and progression are as striking as what they manage to show. Statistical tools come at a price: they provide a partial view not a full summary of attainment and progression, however they may be conceptualised at a particular moment in time. The stronger the belief in rightness of the information they contain, the more likely that very confidence is to be misplaced.

We need more robust challenges from across different contexts of use over what the potential of statistical and administrative data are, their limits as well as their strengths. There certainly is a value and a virtue in numerical data well constructed and deployed with care. But current uses of data in education look increasingly careless. Over enthusiastic adoption and misapplication of a small number of statistical techniques for the purposes of monitoring and control are preventing many aspects of good teaching from being understood or fully recognised. Statistical techniques badly need to be re-appropriated and redirected to help education flourish in other ways.

References

Banks, O. (1955)*Parity and prestige in English secondary education : a study in educational sociology.* London: Routledge

Biesta, G. (2009) Good education in an age of measurement: on the need to reconnect with the question of purpose in education, *Educational Assessment, Evaluation and Accountability*, 21:33–46

Bishop, A.S. 1971. The Rise of a Central Authority for English Education. Cambridge: Cambridge University Press.

British Parliamentary Papers. (1861) XXI *Newcastle Commission. Report of the Commissioners Appointed to Inquire into the State of Popular Education in England, Volume 1-6*. London: Eyre and Spottiswoode for Her Majesty's Stationery Office

British Parliamentary Papers (1862) VIII 473 Minute of the Committee of the Privy Council on Education establishing a Revised Code of Regulations (Reprinted with proposed Alterations, and marked to show which Articles are retained from the Code of 1860) (7)

British Parliamentary Papers (1865) VI *Report from the Select Committee on the Constitution of the Committee of Council on Education (The Pakington Committee).* London, Her Majesty's Stationery Office.

Broadfoot, P. & Black, P. (2004) Redefining assessment? The first ten years of *Assessment in Education, Assessment in Education,* 11:1, pp7-27

Burt, C. L. (1962) Mental and scholastic tests (4th ed) London: Staples Press

Commissioners of National Education in Ireland (1835) *Third Book of Lessons for the Use of Schools*. Dublin: Thom

Committee of Council on Education, Annual Reports, 1859 – 1872-3,

Cowen, R. (2002) Moments of time: a comparative note, *History of Education: Journal of the History of Education Society*, 31:5, 413-424,

Davis Cross, M. (2013). Rethinking epistemic communities twenty years later. *Review of International Studies*, 39, pp 137-160

Dempster, J.J.B. (1954) *Selection for Secondary Education*. London: Methuen

Department for Education (2010) *The importance of teaching: the schools white paper 2010*. London: Her Majesty's Stationery Office.

Fitch, (1859) Examination Schemes, and their Incidental Effects on Public Education. In G. Hastings (ed), *Transactions of the National Association for the Promotion of Social Science,* 1859, P220

Galbraith, G. R. (1997*) Reading lives: reconstructing childhood, books, and schools in Britain, 1870-1920.* New York : St Martin's Press

Goldstein, H. and Leckie, G. (2008), School league tables: what can they really tell us?. *Significance*, 5: 67–69.

Gordon, P. and Lawton, D. (1978) *Curriculum change in the nineteenth and twentieth centuries* London : Hodder and Stoughton

Green, A. (1990*) Education and state formation: the rise of education systems in England, France and the USA.* London : Macmillan

Haas, P. (1992) Introduction: epistemic communities and international policy coordination. *International Organization*, 46, pp 1-35

Hansard, 1861-1862 downloaded from http://hansard.millbanksystems.com/ 24th March 2014

Heim, A. (1957) Letters to the Editor, *Education* Vol 37, p13

Kay Shuttleworth, J. (1861) *Letter to Earl Granville, K. G., on the Revised Code of Regulations contained in the Minute of the Committee of Council on Education Dated July 29th 1861.* London: Smith, Elder and Co

Lancaster, J. (1821) *The Lancasterian System of Education, with Improvements*. Baltimore: Wm Ogden Niles

Laurie, J. S. (1866) *Laurie's Graduated Series of Reading Lesson Books.* London: Longmans, Green, & Co.

Landahl, J. and Lundahl, C. (2013) (Mis-)Trust in Numbers: shape shifting and directions in the modern history of data in Swedish educational reform. In M. Lawn. (Ed) *The Rise of Data in Education Systems: collection, visualization and use.* Oxford: Symposium Books

Lawn, M. and Ozga, J. (2009) 'The Sleep of Reason breeds Monsters' *CES Briefing No 52*, Edinburgh, Centre for Educational Sociology.

Layton, D. (1973) *Science for the people : the origins of the school science curriculum in England*. London: Allen and Unwin.

Leckie, G. and Goldstein, H. (2009), The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172: 835–851.

Lefstein, A. (2013) The regulation of teaching as symbolic politics: rituals of order, blame and redemption, *Discourse: Studies in the Cultural Politics of Education*, 34:5, pp 643-659

Lefstein, A. and Snell, J. (2014) *Better than Best Practice: Developing Teaching and Learning Through Dialogue.* Abingdon: Routledge

Luke, Allan (2010) Will the Australian curriculum up the intellectual ante in primary classrooms? *Curriculum Perspectives (Journal Edition)*, *30*(3) pp 59-64.

Lowe (1848) The Australian Journal, February 14, 1848. p2 Letter dated Sydney, January

Maton, K. 2013. *Knowledge & Knowers*. Abingdon: Routledge

Michael, I. (1987) *The Teaching of English: from the sixteenth century to 1870*. Cambridge: Cambridge University Press

Middleton, J. (2004) The overpressure epidemic of 1884 and the culture of nineteenth-century schooling, History of Education, 33: 4, 419-435

Montgomery, R. J. (1965) Title: *Examinations : an account of their evolution as administrative devices in England.* London: Longmans.

Morrish, I. (1970) Education since 1800. London: Allen & Unwin.

Moss, G. (2009) The politics of literacy in the context of large-scale education reform, *Research Papers in Education* 24:2, pp155-174

M.U.R. (1861) The Revised Code. *Papers for the School Master*, Volume 11. , Dec 1861, pp 233-236

OECD (2011), *Lessons from PISA for the United States, Strong Performers and Successful Reformers in Education*, Paris: OECD Publishing. *http://dx.doi.org/10.1787/9789264096660-en*

OECD (2013) *What Makes Schools Successful? Resources, Policies and Practices – Volume IV*. Paris: OECD Publishing

Sahlberg, P. (2009). Learning first: School accountability for a sustainable society. In J-C Couture, K.D. Gariepy, & B. Spencer (Eds.), *Educational accountability: Professional voices from the field*. Rotterdam: Sense Publishers,pp1-22

Sahlberg, P. (2011). The Fourth Way of Finland. *Journal of Educational Change*, 12(2), 173-184
Schriewer, J. 2014. "Neither orthodoxy nor randomness: differing logics of conducting comparative and international studies in education" *Comparative Education*, 50(1): 84-101

Simon, B. (1953) *Intelligence testing and the comprehensive school.* London: Lawrence & Wishart

Stigler, S. M. with Kruskal, W.H. (1999) Normative Terminology. In Stigler, S.M. (1999) *Statistics on the Table: The History of Statistical Concepts and Methods.* Cambridge, Massachusetts: Harvard University Press

Sutherland, G.(1984) *Ability, Merit and Measurement: Mental Testing and English Education, 1880-1940.* Oxford: Clarendon Press

Sylvester, D.W. (1974) *Robert Lowe and Education*, Cambridge; Cambridge University Press

Thomson, G. H (1921) The Northumberland Mental Tests. *British Journal of Psychology. General Section*, 12:3, pp 201-222

Thomson, G. H. (1924) *Instinct, intelligence and character : an educational psychology.* London: Allen & Unwin.

Tilleard, J. (1860) *On Elementary School Books*. London: Longmans, Brown, Green, Longmans and Roberts.

Vernon, P. E. (1960) *Intelligence and attainment tests*, London: University of London Press.

Vincent, D. (1997) The Domestic and Official Curriculum in nineteenth Century England. M. Hilton, M. Styles, V. Watson (Eds) *Opening the Nursery Door: Reading, Writing, and Childhood, 1600-1900*. Abingdon: Routledge

Waldow, F. (2001) The Suggestive Power of Numbers. Some Remarks on the Problem of the Accuracy of Quantitative Indicators in Comparative Historical Research, *Historical Social Research, Vol. 26 — 2001 — No. 4, 125-140*

 White, G.(1858) Teaching Reading by the Simultaneous Method , *The Educational Record*, Vol 2 , p230

Wiseman, S. (1961.) *Examinations and English education*. Manchester: Manchester U.P

Wooldridge, A. (1994) *Measuring the mind : education and psychology in England, c.1860-c.1990* Cambridge: Cambridge UP

Yates, A. and Pidgeon, D. (1957) *Admission to grammar schools : third interim report on the allocation of primary school leavers to courses of secondary education*. London: NFER/Newnes