

Testing, testing: investigating the epistemic potential of systemic tests

Ursula Hoadley & Johan Muller
University of Cape Town

Introduction

Tests and testing have come under almost universal attack recently from critical educational analysts of a sociological persuasion. The journals on educational policy teem with papers that put tests and testing in the dock for carrying managerialist ideologies, fostering unhealthy competition, and inadvertently promoting deformative pedagogies like ‘teaching to the test’. The impression created by some of this work is that tests are a form of audit and control thought up by regulatory agencies that work on education from the outside, mould it in a particular way, and leave the participants – teachers, learners and their parents – worse off than they were before. This ‘externalist’ critique cannot easily envisage that testing can have a positive and educational role to play, although there are exceptions. Testing is, in this critique, solely a technology of control, and teachers and learners alike are, curiously, deprived of any agency in the process.

This barrage has, perhaps not unexpectedly, had little to no impact on the steady growth and spread of large scale testing programmes. Whoever the critics think they are addressing their concerns to, certainly don’t seem to be listening, confirming the critics’ worst fears that testing has nothing to do with ‘proper’ education and everything to do with accountability, regulation and control. Of course, there is another literature on testing, from the educational psychologists who design these tests, where another picture of testing emerges, and where it is clear the concerns of the critics are also being roundly ignored. This literature emphasises the positive benefits of testing, mainly by stressing the improvement in outcome scores, on the assumption that test scores reflect learner knowledge, or at least mastery of the curriculum (see Phelps 2012 for a review). But in the opposite way, this work also black boxes the classroom, the teacher, and the learners.

The argument to be made in this paper is not that the critics do not have valid points to make. Most of the shortcomings and limitations they point to have some foundation. But the critique is too totalising, and does not admit of asking whether testing can have educational benefits – that is, asking about the *internal educational goods* of testing and assessment, and asking whether there are some contexts where testing can advance the educational enterprise. We will proceed by outlining some of the main lines of critique and defence. Then we will describe the late emergence of systemic testing in SA, and report on some data that brings to the fore the ‘ambiguity’ of testing for

teachers, as Parlo Singh has most recently described it. Tests can help the cause of a decent schooling by supporting teachers and principals, we will conclude, but not uniformly and not unambiguously, and probably not in all situations.

Contesting times

Consider the following:

‘One hundred years’ of *evidence suggests that testing increases achievement ...* Effects are moderately to strongly positive for quantitative studies and are very strongly positive for survey and qualitative studies’ (Phelps, 2012, 39), this from a meta-study of 245 studies conducted over the span of a century in more than 30 countries. Now consider this:

‘Notwithstanding the *very patchy evidence that such programmes are associated with any significant raising of student achievement ...* they continue to be grasped by political leaders and policy makers ... (Polesel et al, 2013, 13: both our emphasis). The gulf between the two sides in the debate here would seem to be unbridgeable. Richard Phelps and others present empirical evidence that seems to show that testing is associated with raised test scores. John Polesel, one of the more measured voices in the discussion, and others, remains sceptical. At least Polesel ends his comprehensive review by calling for more empirical demonstration about the effects of testing, but the question arises – what would convince him they could have any positive effect? Phelps for one doubts that the critics entertain the possibility at all, and dismisses the sociological critics as merely dabbling in ‘popular complaints’ and looking solely for negative effects (It should be said that while this is undoubtedly true in many cases, it is not true in Polesel’s case). Indeed, it sometimes seems as if the critics doubt that tests measure academic achievement at all, that tests measure some impoverished proxy of compliance with a deformative curriculum, measured by even more deformative instruments – at worst, the dreaded multiple choice.

The litany of ‘popular complaints’ is depressingly familiar by now. Their familiarity, of course, does not invalidate them. There is a set of critical comments on testing as a new tool of regulation and governance, a branch of ‘New Public Management’ and ‘neo-liberalism’ which amounts to a critique of the new governmentality, and although significant, will not be further pursued here (see Lingard, 2011, Au, 2008, also Yates, 2013. A particularly measured account is provided by Hopmann, 2008). The focus in this paper is on the pedagogic potential of testing. In regard to pedagogic impact, a useful distinction is made by Jones (2007), who distinguishes between the *intended* outcomes of testing – higher achievement scores, which he does not dispute, and the *unintended* outcomes, a distinction also made by Smith & Kovaks (2012) and Lingard & Sellar (2014). It is these ‘perverse effects’ that seem most concerning, and demand consideration. They include teaching to the test (curriculum ‘narrowing’), omission of valuable curriculum topics because more time is spent on reading, writing and maths than before (if only our teachers would do that ...). ‘Narrowing’ is sometimes regarded

as ‘gaming’ the system, and somehow illegitimate. Yet it is surely the natural consequence of a greater visibility and focus on coverage of these specific so-called ‘gateway’ competences, which forces a slackening of pacing to accommodate all learners, which in turn takes up more time. The pedagogic lesson here is that changing the pacing of pedagogy, if not explicitly accommodated by time slots provided elsewhere – a longer school day, or week, or after-school sessions – will inevitably impinge on curriculum time within the conventional allotment, necessarily displacing something. In other words, this is not a ‘perverse effect’ so much as a consequence of bad curriculum planning and time management.

More serious are charges that the ‘narrowing’ effects of testing can limit not only the breadth of learning on offer (Lobascher, 2011; Au, 2008; Barrett, 2008; Hirsch, 2013 amongst others) but its depth too, and that testing can promote a surface and atomistic approach to learning (Lobascher, 2011), that stunts creativity (Madaus et al, 2009) but also hypothetical thinking and problem solving (Tanner, 2013). Tanner’s is the most serious charge: it would be unfortunate indeed if tests produced robotic information processors without the cognitive connective tissue to turn the information into knowledge. Fortunately, the evidence for this is quite thin, and Tanner is rather speculative in this regard. At best, the evidence is based on what teachers report might be the case, or on anecdotal comments by university professors on the declining ‘quality’ of school leavers entering university.

A final critique is that testing, as part of the larger ‘audit culture’, threatens teacher professionalism by driving out ‘a humane relationship to knowledge’ (Beck & Young, 2005, 184) a relationship lying at the heart of all genuine professionalism. In his most recent book, Young (Young & Lambert, 2014) sharpens the critique to argue for a ‘less assessment oriented approach’ (42), which will allow teachers as curriculum specialists to teach a ‘knowledge oriented’ curriculum based on the trust of the public and government, who is ‘always going to overemphasise assessment and league tables’ (ibid). There is something a little ironic about this assessment of assessment. The villain of the piece, admittedly more emphatically in the earlier paper than the later book, is clearly the government. But might it not be the other way around? Might it not be the case that it is the teachers who have misplaced the ‘knowledge oriented curriculum’, and might not the consequent loss of trust in these expertise-disavowing professionals have at least something to do with the ‘audit culture’ and the rise of testing? Indeed, might it not be incumbent on government, in such knowledge-misplacing times, to check whether their teacher professionals were doing what, in a more innocent and responsible time, they were routinely trusted to do? However that might be, we certainly also agree wholeheartedly with Young’s conclusion that looks forward to a time when tests and exams *support* what teachers do rather than *determining* it or driving it (ibid, 43). Our task in this paper is to inquire into how such ‘support’ could actually work, and whether teachers are receptive to such scaffolded support.

Then there are the purported pedagogic benefits of testing. Polesel and his co-workers usefully summarise these as follows. Testing provides:

- More reliable grading than teachers (Phelps, 2006);
- Clearer information to learners (and their parents) which areas of work they have mastered and which not (Sloane & Kelley, 2003);
- Clearer signals to schools about what is important in the curriculum (Amreich & Berliner, 2002);
- Greater curriculum consistency across schools (Clarke et al, 2003; Crocker, 2004; Jones, 2007; Santiago et al, 2011). This is a critical point, because it forms the basis for a social justice argument for testing, as found in Australia for example, rather derisively referred to by Lingard et al (2012) as 'gap talk'.

These are all useful pointers, but, in time honoured school effectiveness style, are offered as stand alone technical solutions, not placed within a conceptual context where they might hang together in some way. This is what we attempt in this paper.

There are a number of observations to be made before we wade further into the melee. The first is that it should be clear that not everybody is talking about the same thing. The ultra high stakes of testing in the USA in the wake of NCLB and RTTT cannot meaningfully be compared to far lower stakes systems in Australia, and SA. In fact, in the Australian case, NAPLAN is by most definitions (see for example see the definition by Baker et al, 2010) a low stakes system, with their *MySchool* website providing a public site for symbolic shaming but otherwise no other material penalties as there are, and brutally so, in the USA. The damage such publicity can cause to what Lingard and Sellar (2013) have called 'reputational capital' should not be minimised though, especially in the context of a still-fragile national culture that prizes the maintenance of social face, which is probably why Polesel et al conclude that it is, after all, to be considered a high stakes system.

Whether or not the Australian system is or is not high stakes is not really material; the larger point has to do with what effects stakes have on the potential of testing. Psychologists involved in testing have for years emphasized that attaching high stakes to testing threatens to eclipse the positive diagnostic potential of tests, and triggers instead compensatory responses that may be pedagogically harmful (see Hattie, 2009 for example). In 2010, the American Economic Policy Institute issues a policy brief authored by 10 authoritative educators convened for the purpose, who concluded: 'Statisticians, psychometrists and economists who have studied the use of test scores for high-stakes teacher evaluation (test results used to reward or sanction teachers)... mostly concur that such use should be pursued only with great caution' (Baker et al, 2010, 7). Ultra high stakes testing is simply a bad idea, and no responsible educator should today defend it. We certainly don't. The point we are making here, though, is that it is reasonable to suppose that the most pernicious effects on pedagogy attributed

to testing may be triggered by the stakes attached, not by the technology of standardised assessment. Yet it is precisely in highly contested melees like this one that these things get run together. The vulnerability of teachers in testing times is heightened, and vulnerable professionals will not always put the learners first. We will return to this point.

The second point is related. There is a tendency in high contest debates to resort to what Terri Seddon called 'big brush-stroke stuff' (Seddon, 2003, 243, cited in Singh et al, 2014). She had the label 'neoliberal' in mind. What this can lead to, in turn, is the tendency to read the rest of the argument off from the first move, not off the evidence, leading to what Parlo Singh and her colleagues, following Stern, call a 'paranoid reading': 'A paranoid reading of current education policies automatically *knows* these policies as the products of a neoliberal market logic' (Singh et al, 2014, 5, quoting Stern). The rest follows.

None of this helps, and it eclipses two features that should be kept in mind when pedagogical features of a system like testing are considered. The first is that *contexts matter*. For example, the persistence of neoliberalism in the USA is intimately tied to the American sense of individualistic liberal democracy, as Labaree (2012) has persuasively argued. It has a particular purchase there it may not have in countries and systems into which it has been imported. Lyn Yates (2013) is particularly attuned to contextual specificities, but hers is the exception, not the rule. As we will see presently below, the SA context has some specific features that require particular attention if we are to avoid getting pulled into the distinction-collapsing melee.

The second is that, from a particular angle, both the proponents and critics pass over the teachers rather quickly, occluding in particular the *possibility of their pedagogic agency* (see also Stern in Singh et al, 2014). The proponents focus all too often on the achievement results as direct causal outcomes of testing – constructing in the process the familiar 'black box' of the pedagogic work of which the scores are a product. Equally, however, we find that the critics too often cast the teachers in the role of passive victims of totalising regulatory control of which testing is an integral part. Although their avowed intent is to provide a critique that will somehow empower the teachers, the 'paranoid' critique all too often forecloses the possibility of pedagogic agency in its aim to put testing in the dock. In this paper, we wish to raise the question explicitly as to what the possibilities for pedagogic agency are in the light of tests and test results.

Two steps forward

There are two recent papers we have found helpful in sharpening our thinking in this regard. The first is Lyn Yates' (2013) recent paper written in response to Daniel Tanner. Yates' account is a nuanced one that does a number of things that take the

discussion around testing a few steps forward. First, in a gentle rebuke to Tanner, she emphasises the importance of national context. In Australia, testing is rhetorically justified on the basis of its role in making visible educational inequities and pointing at where state funding should be directed, and she concedes that this visibility is indeed salutary. She is also happy to concede a useful role for testing: 'Assessment and testing has a role in curriculum, especially formative and diagnostic; and the construction of appropriate assessments as part of the curriculum process is one way of making concrete and visible to students what is important in a curriculum plan' (42). We agree completely. She also concedes that 'it is reasonable to have some points at which the quality is assessed of what children gain access to and the distributional inequalities that are evident in the outcomes' (44). Agreed.

In other words, although she accords a place for testing in the New Public Management and 'audit culture', she is more sanguine about testing prospects than most of her critical sociological stable mates. It is the collateral effects that trouble her, and these are twofold. The first is that tests unreasonably elevate expectations about what schools can be expected to do to correct social inequities. We agree there too: testing is not going to interrupt class reproduction. She goes on to outline a usually obscured reason for this, pointing not at the machinations of either regulatory forces or the interests of capital, but at the already privileged middle class parents. Testing results not only make visible inequities, but ramp up what she calls 'competitive anxiety' about the potential loss of social position and standing. In other words, achievement publicity not only facilitates compensatory government interventions, but drives up also the lengths middle class parents will go to in order to maintain their competitive advantage and that of their offspring. They can do this because they have the wherewithal to do it.

This is a chastening account, but a not wholly pessimistic one. Testing will not interrupt class privilege. But can it not at least do something a little more modest? This is the question asked by Parlo Singh and her colleagues in a recent paper with which this paper has many affinities. Singh roundly concedes the points of critique discussed above, but asks what we then do, as engaged scholars, to push beyond the feeling of powerlessness this critique avalanche can engender, and help the vulnerable and apprehensive teachers to explore 'possibilities of working differently with assessment data' from, to use Stern's words, 'somewhere more immediate' (5), somewhere beyond the 'paranoid reading'. That 'somewhere' turns out to be the classroom, and the 'working differently' building educative partnerships between university researchers, 'school based researchers' and teachers. She calls this a 'dilemmatic space', 'where participants held paradoxical and ambivalent feelings towards standardised testing regimes' (5), but were helped to work through them. The 'work' in this space consisted in working first to build trust with the unnerved teachers, helping them to look at the test results more analytically (our description), pushing beyond anxiety to co-construct strategies from the test data. Revealingly, in order to do this, the teachers had to learn,

in the words of Webb & Gulson, to 'face themselves differently' (cites Singh et al, 2014, 10), that is to say, to simultaneously 'engage' and 'distance' themselves from 'individual students and their own teaching practice' (15).

Psychologists with long practice of working with tests results will call this 'feedback' (Hattie & Timperley, 2007; Hattie, 2009) and feedback is a central part of the New Zealand strategy regarding assessment. This description probably under-estimates what is at stake here. As the American high stakes system crudely underlines, testing puts not only the learners under the spotlight, but the teachers too. Unsurprisingly, they usually feel exposed, vulnerable and defensive – as most would when brought to the mark. This is greatly enhanced when teachers don't know how to 'read' the test results, especially when they know at least that the news is not good. Making the test results legible, and helping to craft doable strategies in response to them is the function of mediation which puts the feedback to work (Archer, 2010), and what Singh et al's partnerships evidently achieved, since the intervention schools' test results improved dramatically, over 40% greater improvement than in non-project schools. This is highly significant.

What then was it that the teachers learnt? They increased their capacity for working with evidence to enhance literacy instruction. Their knowledge about how learners learn to read and think was enhanced. They were more effective in selecting teaching and learning strategies. And all of this boosted their confidence; they were convinced that the achievement gains were as a result of changes they had introduced – empowerment by any other name. Yet this was to be expected: expert interventions improve pedagogic strategies. In our paper we want to take a step back and ask: what was it the teachers read in the test results that fed their pedagogy? What curricular information did the test results hold for teachers when they learnt to decode the results? How did the 'diagnostic' potential of test results manifest itself? The work done by Singh et al persuade us that test results can, when correctly mediated, abet rather than merely distort classroom pedagogy. But how?

The context of testing in South Africa

South Africa has a relatively recent history of standardized testing throughout the schooling system. Until recently, a single, high stakes exit examination was administered in the final year of schooling, and no external testing was conducted in the rest of the system. During the course of the 2000s two standardized tests became entrenched in the system – one at the national level, the Annual National Assessment (ANAs), the other a Western Cape provincial testing regime called the Western Cape Systemic Evaluation (WCSE) (referred to in the data below as 'the systemic'; the 'WCED'; and the 'LITNUM').

The WCSE was implemented as a provincial programme in the Western Cape for the first time in 2002 at the Grade 3 level, followed by Grade 6 in 2003. At first it was a sample-based test, and administered alternately in Grade 3 and Grade 6 every two years. In 2010 the test was extended to Grade 9, and for the last four years the test has been administered annually in Language and Mathematics to all Grade 3, 6 and 9 learners in all public and most independent schools in the province. There is no public reporting of test results by school. Reports are sent to schools that provide them with their mean score against the mean score for the district and the province. In 2012 results were reported to schools at the learner level. The tests are externally administered and moderated. There are no sanctions for poor performance in the test, however, particularly low results do lead to increased monitoring and support. Rewards for improved results include financial payments to schools.

The Annual National Assessments, the first programme of national population testing across a range of grades, was instituted in 2011. The ANAs were administered to all learners in grades 1 to 6 and Grade 9 in mathematics and language. The tests are administered and moderated internally at the school level. In addition, a verification process conducted by external agents in a sample of schools determines the degree (if any) to which school-based scoring differs from objective scorers. The programme has been fully implemented, but given its newness, there are serious capacity issues both in the design and the administration of the tests. Issues in design and construction limit the usability of the results. In particular, problems pertain to the lack of benchmarking of test items to a predetermined (curricula or other) standard of competency. Also, test construction across the different cycles of tests is problematic and there is currently no statistical foundation for making comparisons between different grades or across test cycles (World Bank, 2013).

The new testing regimes, the most prominent being the ANAs and the WCSE tests, are low stakes with respect to accountability functions and immature with respect to diagnostic purposes (especially the ANAs). Though developing, both are currently hampered in different ways by inexperience in large-scale testing regimes across the system. Further, there are a number of interrelated features of the context of testing that we want to highlight briefly here in relation to our discussion below.

The first is that the South African schooling system is defined by extremely low learner performance, both on the national tests (ANA and WCSE) and on cross-national tests such as TIMSS, PIRLS and SACMEQ¹. These tests cumulatively and consistently show that across the schooling system learners are not reading, writing or calculating at the appropriate grade level. Research has identified a dual system of primary schooling in South Africa, which produces bimodal patterns of performance (Van der Berg and

¹ SACMEQ – Southern and Eastern African Consortium for Monitoring Educational Quality, TIMSS – Trends in International Mathematics and Science Study, PIRLS – Progress in International Reading and Literacy Study.

Louw, 2006; Taylor & Yu, 2009; Fleish, 2008). 25% of the system performs at satisfactory levels, whilst 75% of the system is failing, with national averages of 30-35% on tests of numeracy and literacy being the norm in this group of schools (Spaull, 2012). The two sub-systems are split along former-department, colour, language, and socio-economic status lines.

Secondly, the inability of teachers to accurately assess their learners on internal assessments is well documented in the South African research literature (van der Berg & Shepherd, 2010; Lam et al, 2011). This relates in turn to historically low levels of teacher selectivity into the system and very low levels of teacher content knowledge (Venkatakrisnan & Spaull, 2014; Taylor and Taylor, 2013). Finally, teachers in South Africa are highly unionized and weakly professionalized (de Clercq, 2013). There is a strong collective culture among teachers and in teaching, underscored by weak individual performance management systems. All of these features combine to contribute to huge inefficiencies in the system. The post-apartheid official curriculum did not help. It's low content specification, outcomes-based approach meant low levels of content signalling, especially for teachers who lacked a knowledge standard given their own schooling, training and teaching contexts.

Given their relative immaturity, the diagnostic and accountability aspects of both tests are still in development, with the WCSE systemic tests having developed some way beyond the ANAs. One of the clearest outcomes of the tests thus far though has been the extent to which they put disparities in educational outcomes between different social groups publicly and politically on display (Yates, 2013). In relation to our data, we consider however, the effects of testing at the level of the school. Given the contextual features of testing sketched above, we were interested in the pedagogic potential and implications of tests in such an environment specifically in relation to the ways in which schools and teachers in contexts of poverty respond to these standardised testing and reporting policy regimes.

Sample and methodology

The sample is drawn from the SPADE project which used the results of the WCSE Language and Mathematics tests for Grade 3 from 2002 to 2008 (the results of four test cycles) to select schools. All South African public schools are divided into poverty quintiles, largely based on levels of income, unemployment and illiteracy within the school's catchment area. Quintile 1 schools are the poorest institutions and quintile 5 the well off public schools. The quintiles dictate the level of equitable funding the school receives from the state. Learners in Quintiles 1, 2 and 3 obtain a much bigger subsidy from the government (R1 010 per learner registered in 2014) compared with learners in quintile 4 who obtain on average half that amount (R505 in 2014). In quintile 5 learners are subsidised at roughly 10% of that of the poorest schools (R174 in 2014). Quintile 4 and 5 schools are expected to supplement their state allocation

through the charging of school fees and fund-raising. Schools selected for the SPADE project were all in quintiles 1 to 3 (fee free). Selected schools had achieved an overall mean for the four-stage period that was at least 5 per cent *above* the overall mean for their former departments, that is, for their socio-economic and cultural peers². This generated a total selection of nine schools – five former HOR schools and four former DET schools. The nine schools represent the *full population* of large primary schools in the lower 3 quintiles achieving above their former-department means in the WCED tests in the Western Cape province, ‘punching above their weight’ as Polesel et al, (2013), following the ACE report, describes it.

Following the initial selection, five more schools were selected which *matched* the demographic profile and former racial department of the nine higher achieving schools but whose averages over the same period were 5 per cent *below* the mean of their former departments. The final sample of 14 schools was made up of 8 former HOR schools and 6 former DET schools. These schools were organized into five matching sets, with one high and one medium performing school, and one lower performing comparator school. There are two things to note about the selection of schools. First, although the sample is small, it is the *total number* of schools that performed above average within our criteria. In other words, few schools are managing to achieve even modestly above expectations (or average for their former department) and thus they constitute a unique sample. Secondly, none of the schools are performing well above the norm for their former department, there were no ‘outstanding’, or even conventionally ‘good’ schools in poor communities to be found in relation to performance. Other analyses in the broader SPADE project have compared the two sets of schools. In this particular analysis, we take the sample as a whole. Given that the schools are managing to perform just below or just above the average, we assume that they are both possibly more responsive to policy reform or opportunities around testing. There were no discernible differences in relation to school and teacher responses found between the above and below average performers.

The analysis is based on 67 interviews conducted over a four-year period in the 14 schools, with the principal, heads of departments and teachers. In addition, data from 14 focus group interviews with Grade 3 teachers was also gathered. All interviews were conducted in English. The home language of all respondents was either isiXhosa or Afrikaans. However, they were all fully competent and comfortably conversant in English except for three teachers who spoke Afrikaans. Although responses to testing from schools leaders (principals, HODs) is likely to be different from teachers (Hardy, 2013), we don’t differentiate between respondents in our analysis, finding a difference

² These are the former race-based education departments of the apartheid regime. In this case DET refers to former black schools, and HOR to former coloured schools. The DET schools remain largely racially and linguistically homogenous, whilst the HOR schools are in the process of changing in relation to both characteristics. The classifications were used for sampling because these former apartheid designations remain the central predictor of schooling outcomes.

in focus and emphasis in response rather than a difference in opinion or fundamental claim around testing in the schools.

The diagnostic potential of tests and the evaluative rules: visibility and the pedagogic device

While quite a number of writers acknowledge the diagnostic potential of tests (recently Yates, 2013 for example), this has rarely been theorised or further explicated. Testing concerns evaluation, and is central in generating the outcomes of schooling. In Bernstein's (1996) terms, testing forms part of the broader notion of 'evaluative rules' which are located in his conceptual model that he termed the 'pedagogic device'. The pedagogic device explains how knowledge is transformed into pedagogic communication. It operates at three levels - respectively, the distributive rules, the recontextualising rules and the evaluative rules. The distributive rules concern what knowledge is distributed to whom. The recontextualising rules show how knowledge is selected and specified in the curriculum. This is what is intended to be recognized and realized by the student. At the level of evaluative rules (pedagogy) the concern is both with what has been recognised and realised by students, and also what has been made available to whom. Bernstein talks about the evaluative rules as 'condensing' the device in this way. What tests potentially offer then is a visible rendering of the underlying order of the schooling that has been provided.

But what tests do, at the level of evaluative rules, is also partly to 'de-condense' the device and it is here that their pedagogic potential rests. Tests show what of the recontextualised knowledge is recognized and made available for realisation by the teacher. At the level of evaluation we can see what was specified at the level of recontextualisation and its consequences. The tests also give a measure of the distributive rule - what was distributed to whom. In this way test results potentially put on display the effects of the distributive and recontextualising rules of schooling.

If at a general level testing renders more visible the underlying order of schooling, more specifically we suggest this visibility is represented in 'codes'. The codes embed a regulative discourse, which refers to the social relations of teaching and learning, to the conduct, character and norms that are expected in the pedagogic process. The codes also embed an instructional discourse - the knowledge, skills and competences to be acquired. Below, we identify three codes which we used to organize the data we gathered around schools' responses to testing. These are a pedagogic code (to do with the selection, sequence and pace of instruction), a transmitter code (to do with the distinctiveness of teachers), and an acquirer code (to do with how learners are pedagogically distinguished). In the analysis we show two things. One is that the diagnostic potential of tests lies in pedagogical visibility. This visibility inheres in codes, that is, its visibility depends upon it being deciphered or de-coded. In the data we find differentiation and individualization as dominant features of the pedagogic,

transmitter and acquirer codes. Differentiation distinguishes between spheres of adequate provision and spheres of need, which in turn allows for the considered distribution of resources in low resource settings, in the interests of improved academic outcomes. The other focus of the analysis is on the ambiguous space that this pedagogic visibility generates. The codes require mediation (that is, knowledgeable de-coding) in order that the critical potential of tests can be realized and the coercive and threatening aspects minimized.

Pedagogic, transmitter & acquirer codes

As indicated above, we identified three ‘codes’ made visible, or potentially visible by testing: a pedagogic code, an acquirer code and a transmitter code. It is through the scrutiny of results, that the schools in the research, to a greater or lesser extent, ‘cracked’ these codes in order to read the evaluative rules and activate the diagnostic potential of the tests. Through looking at instructional dimensions, and categories of learners and of teachers, schools used tests to enhance the pedagogical visibility of what went on in their instructional programmes and the nature of their regulative discourse. At the broad level, patterns in performance were identified by schools within a cohort at a single time, as well as across cohorts. For example, looking across cohorts of grade 3, the principal at School A remarked, “Numeracy for grade 3 rose by 19%; Language dropped. We are seeing a steady upward on WCED. We’ve had a strong focus on maths...”. And the principal at School 1 commented on between different grade cohorts: “Grade 6 is a problem - something is happening between grade 3 and 6. There is also a gap, between grade 3 and 4, and this is also the biggest down jump [from grade 3 to grade 4]”. How instruction, teachers and learners were ‘read’ through considering the tests is discussed in detail below in relation to the pedagogic code, acquirer code and transmitter code.

The pedagogic code

The pedagogic code embedded in the test makes visible features of the instructional discourse of the school, most crucially, following Bernstein, the selection, sequencing and pacing rules. In relation to selection (what content is taught), tests function in two ways. They drive curriculum coverage and indicate areas that require targeted agency:

“It is easier to teach the other grades [ones not being tested]. These grades [where the test is administered] you have to teach the curriculum” [School 4, Teacher JON].

“Teachers try to complete the curriculum before the final assessment/exam but [it] differs between classes. It’s easier for those not involved in systemic tests not to finish. Also in relation to ANAs. Before this, because [it] takes long for learners to learn, [it] took longer” [School 7, Teacher EUR].

It make explicit key curriculum contents to be covered at both a specific and a broad level: “what we realized is that learners *have* to learn to read, write and calculate” [School A, Teacher HAR].

The *sequencing rule* refers to the ordering of content in relation to its content complexity. It is about the developmental progression of content within and across grades. In relation to the sequencing rule, the tests are an important signal of coverage and complexity. At School D the principal refers directly to the issue of ordering and level of curriculum content: “Do teachers know what to do? Next, what is happening in the classroom? How do learners cope with assessment tasks? The challenge is for teachers to get classrooms up to the level”.

Given teachers’ weakness in internally assessing their learners (van der Berg et al, 2006), the signalling of the sequencing rule (expected order and level of complexity) provides an indicator of expected level where teacher judgment has proven too variable. Below are two examples where HODs identify the problem of internal assessment elicited by the tests:

‘We know there is a problem because we do well on our internal tests but not on the systemic tests’ [HOD, School 9]

‘Our own exams and ANAs [internally assessed] are good. The WCED test [externally assessed] not so good. Numeracy for grade 3 rose by 19%; Language dropped. We are seeing a steady upward on WCED. We’ve had a strong focus on maths now...’ [HOD, School A]

Comments relating to ‘level’ and complexity of content were found across the interviews. In the data there are copious examples of ways in which teachers have attempted to deal with the expected complexity of the content in response to signalling from the test.

While selection and the sequencing rules are critical pedagogic signals derived from the test regarding what is taught and at what level, it is the *pacing rule* (the time taken to acquire the sequencing rules) that is the focus of most discourse around the test. Across the interviews the pace of the curriculum and of instruction is referred to. Where students have not grasped the sequencing rules, Bernstein (1996:65) argues that there are three possible strategies to address this. The first is to maintain pace, but to reduce the quality and quantity of instruction. The second entails the introduction of a repair system, which generally involves extending time sequentially or employing additional specialists or extra lessons, which extends it in parallel. The third strategy involves a relaxation of the pace so that the student is given time to meet the requirements of the sequencing rules. In our data we see evidence of strategy

two and three in response to the tests. Here a teacher explains how the tests have led to a relaxation of pace in certain topic areas: “We use the tests to identify specific areas for focus. For example, in Grade 1 learners need to go up to number 34. They need to count forwards and backwards. We can develop a strategy and *then stick to this if the children aren't there yet. We need to stay back until the children can do it* [School 3, Teacher MON]”.

13 of the 14 schools have the second strategy in place, a repair system that has been instituted in the school as a result of testing. In six of the schools additional personnel were specifically hired to assist with instruction in maths and reading. In all schools the repair strategy entailed *the expansion of instructional time* (mornings, afternoons, weekends or holidays). One of the most trenchant critiques of testing in the literature is around curriculum narrowing and the atomization of knowledge in the classroom. Bernstein's first response cited above could be read in relation to this. Narrowing entails a reduction in the quality and quantity of instruction. But as a repair strategy, narrowing can also be read as a means to achieving depth. What emerged from our data were ways to extend the school day in order to accommodate the curriculum, and focus on particular areas of weakness and for those who have fallen behind in the sequencing rules. For example, an HOD at School E describes: “We have a reading program in [the] morning. I supervise it. We are trying to get a variety of texts and focus on techniques for reading”, and at School 7: “We had a foreign student do workshops. We had R28.000 in our budget - we hired an extra student for a reading project. In the first term we had afternoon classes, which added an extra hour to the day. We streamed learners. I'm not sure if it all worked” [HOD].

Strategies two and three involve a process of *differentiation*. One of the central aims of such differentiation is to extend and distribute pedagogic resources more optimally, including the resource of instructional time. The visibility of the pedagogic code assists in judgments of how this resource should be allocated, and in relation to what curriculum areas.

The transmitter code

The transmitter code embedded by the test makes visible distinctions between teachers at both an instructional and regulative level. In relation to instructional aspects, tests allow for differentiation between transmitters on the basis of expertise or specialization. And in relation to regulative discourse, professional commitment or motivation is marked out.

In the sample, schools read the tests at the instructional level to make more visible the specialist requirements of teaching at this level. Teachers are compared in relation to results achieved on the test by their learners. At School A, the principal attempts to extract examples of good practice from results: “One teacher got 70% ...

what do you do to get this result? We don't have time to share but it will be useful to see what she is doing”.

In School 1 the principal talked explicitly about ‘identifying teacher strengths - who is good in maths and languages’. And another principal indicated how this information which was absent in the past, has led to a greater sense of areas of performance and teacher placement:

‘Every term, we talk about learner performance in the SMT [school management team]. We get [mark] schedules, analyze them, compare terms. The HODs go down and discuss in phases. In the first term they talk about the systemic tests. They focus on specific areas if there is bad performance, and how to improve. But we didn't know strengths then. *So now [we] have a sense, so we can move teachers around and address areas that need attention.*’ [Principal, School B].

In School 2, they have a single specialist maths teacher for all classes at the Grade 3 level: “This year we began an intervention because of the failure rate. I am teaching maths for all the grade 3’s ... we are trying. Maybe instead of 20%, we’ll get 30%. I was getting the highest percent. I was getting 75%, 50%.” [School 2, Teacher NQO]. One HOD draws a general lesson: “Last week, I met with principals around here. We need to capacitate teachers. We need more subject teaching from early on, not just generalists” [School 3].

At the regulative level, the testing marks out expected professional compartments which are perceived to be having an effect:

Discipline has really affected our performance, both teachers and learners. We are trying to set new tone, new culture of teaching and learning. Learners didn't know difference between home and school. Now it’s a place of learning. Teachers also – school is a place to teach, not for chatting. *The tests constantly remind them what is expected. They [the tests] make attendance and punctuality important* [Principal School B].

At School C the teachers refer to the way in which the macro-pacing strategy devised by the school principal has ruled out established practices of extended communal lunches shared by teachers: “At school meetings, the principal talks about the ANAs, teachers must be in their classes. The principal worked out how many hours till the rest of the year. *There is no time for a restaurant in class*” [Teacher SMA].

Teachers’ hard work, determination and commitment in relation to improvement on the tests is also emphasized across the interviews. In the last example above, the shift towards a stronger performance-based, individualized regulative order, driven in part

by testing, is evident. In the other examples, it is also clear that a greater differentiation of teachers inheres in the transmitter code of the test – who is strong, who is weak, in what curriculum areas. Further, the specialist requirements of the instructional programme are indexed by test results in different subjects and content areas.

What the transmitter code of the test brings to the fore is an *enhanced opportunity to distribute resources in a low resource context*. By analysing test results, teachers are differentiated and distributed with the aim of optimizing academic performance. This is a significant shift from the collectivized teacher cultures that dominate in low resource schools such as those described in past research (de Clercq, 2013; NEEDU, 2012; Hoadley, 2005).

The acquirer code

The acquirer code refers to the ways in which learners are reflected in the tests, and this crucially entails differentiation of learners on the basis of performance on the test. Students are individualized academically. Schools analyse test scores in order to decipher the acquirer code. Repeated reference to the scrutiny of learner test score data was made across the data. For example, at School 3, the principal explains:

In phase teams, we analyse results in the phase each term. We design strategies for improving results and set targets. Also we make a list of learners who you are going to give extra support to. And you have to then monitor these learners' progress -- like a catch-up program – what effect is it having?

Differentiation of learners in the schools gives rise to extensive systems of repair, shown in the discussion of the pedagogic code above, primarily through extended time and the recruitment of additional personnel. It also gives rise to streaming, a practice hitherto shunned in these schools. In all 14 schools streaming practices were evident. These were generally implemented in the extended instructional time periods that formed part of the repair system. In two schools, School 1 and School E, however, the principals instituted a streaming programme within the mainstream educational programme. From Grade 1 the schools identified the most promising learners and matched them with the strongest teachers, who then followed the cohort up through the grades. From one point of view this seems like a perverse effect that discriminates against less able learners. From another, it is simply trying to do the best in a context of limited resources.

Differentiation occurs both downwards for remediation, but also upwards for the extension of those learners who show academic promise. At School 2 the principal explains how he nurtured the top maths performers: “This year we identified the

best learners in maths on the systemic [test]. We put together a vibrant chess team. And they enter in all the competitions..." And at School 7: "It's now on the timetable. We have a didactic period on the timetable; Monday - Language; Wednesday - Math. With struggling learners, we can also use strong learners to help. From the last period onwards, till 3:30pm" [HOD].

The acquirer code, like the transmitter code, enables the distribution of pedagogic resources. Test results make possible, in the last example, the *identification* of 'strong learners' to assist the weaker learners. Learners are recognized in relation to their *difference*, rather than a communalized grade cohort. While the perils of labelling may well emerge with such differentiation, reading the code clearly offers the potential for academic differentiation and pedagogic remediation and extension that does not reduce all students to a common denominator with respect to the instructional offering available.

Ambiguous space

Across the interviews, there were surprisingly few negative comments about the tests, possibly because the low stakes attached to the test removed any incipient threat to the teachers, though no reasons were explicitly offered. Where negative feelings were indeed expressed, these related to three aspects. The first was a lack of understanding around the meaning of test results, and especially shifts in results. In response to why test results had gone up or down, the principal at School 6 said: "We go up and down; we can't explain difference in performance from year to year. A difference sometimes of nearly 30%" [School 6 Principal].

The second form of negative response was related to the increased pressure placed on schools and teachers by the tests. In all these responses, however, there was ambivalence. In the following extract, the teacher equivocates in her response between the test as a negative surveillance mechanism and the test as a positive means for improvement:

To me, I don't know why we have the test. To catch out a teacher? We are all working. In a way, it's good though. You don't want to be down there, you want to work harder. You don't want to feel out...For some teachers it is very stressful. The principal, at all times says we need to work. That's where some teachers say I want to teach grade 2 and grade 1 now [levels at which the systemic tests aren't conducted]...[School A, Teacher ARE].

The principal from School D expresses confidence in testing at a system level, but ambivalence at the level of the school, particularly about the intervening effects of home background in test results, and the demoralising effects of disappointing results:

‘The low morale of teachers negatively affects them, but they want to do their best. They feel that with the help of parental support they could do better. Where there are exceptional cases of support, learners do better. For the leadership at school - personally it’s a challenge when faced with all those issues and one feels one doesn’t get results. It should not only be about figures, but also how happy learners are. Still, the system is working’.

The third form of negative response is related to issues of reputation, and a form of professional shame provoked by the tests, much like teacher responses elsewhere (see for example Polesel et al, 2013). Even though schools are not formally compared through league tables, the comparative potential of a standardized test is felt in some of the schools. In School 9, a teacher points out the lack of fairness in placing schools with very different histories on the same measure, ‘model C’ schools referring to former ‘white’ schools that were well resourced: “I don’t like it [the test]. They are comparing us with the ex-model C schools. Our children don’t have books at home” [Teacher MBE].

And at School 1, Teacher CAM expresses the anxiety of being evaluated through the tests: “During a typical grade level meeting, we talk about certain aspects to focus on, preparing for the systemic [test]. We are stressed. Our names go out with the marks. We need to look good and do better in class”.

The principal at School 3 draws attention to the reputational implications of positive results:

“In 2009 we were clapping hands. We got R15000 and a certificate. We were excited, smiling. Other principals were looking at me. Then in 2011 writing was a problem and time was a problem”.

At School D, the principal speaks about reputation in relation to fluctuations in test results:

‘We have always been good, climbing [the] ladder of success. But we feel down about 2011. For teachers this can be very demoralizing. The maths teachers are very good; there is a degreed teacher there. And they are very down from the 2011 results. From 2008 to 2012 we were the best coloured school. In 2009 we got an award, R30,000 for [performance on the] systemic [test]. Everyone is watching for our 2012 results’ [Principal, School D]

The tests provide teachers and schools, probably for the first time for many, an occasion for their practice to be seen through the eyes of others, to ‘face themselves

differently' (Webb & Gulson, cited in Singh et al, 2014). This produces a form of professional shame in cases of poor performance; pride in the context of improvement. What is interesting about this reputational effect in this data is the way in which schools and teachers reflect on results in relation to a particular category of school. In the comment from Principal D above, he heads the 'best coloured school'. In other interviews the school is 'in the district', or the comparison is 'amongst no fee schools'. Rather than a discourse of disempowerment, schools also respond positively to equity type categories, and understand test scores as relative. The evaluative rules are about discrimination, and the schools have clearly begun to read the codes provided by the test and discriminate between different kinds of instruction, teachers and students within broader designated equity 'types'.

Concluding remarks

Rather than a sense of 'hopelessness, despair and obstructed agency' in the face of testing identified in the literature (Singh et al, 2014), the overwhelming sense from the interviews, from many teachers and all the principals, is the sense of being rendered visible and *being noticed*. The tests produce a differentiation of instruction, teachers and learners at the micro level. At a more macro level differentiation means high poverty schools *not* seeing themselves and being seen as all of a disempowered piece, indistinguishable in the 75% part of the bimodal distribution. What the tests introduce are subtle distinctions between schools where a collective resignation to failure is potentially interrupted.

The *lack of specificity* in the interviews with respect to instruction suggests that visibility of the pedagogic code in the schools in the study is still in a *relatively weak form*. However, there is a greater fundamental awareness of content selection, the sequencing rule (progression) and the pacing rule. We are not making strong claims about the knowledge signalling potential of tests at their current stage of design and implementation. Nor are we claiming strong diagnostic skills on the part of schools and teachers (see also Sapire & Herholdt, 2014). Rather we are drawing attention to the positive way in which differentiation effects are produced in schools, particularly in relation to the capacity fostered for the allocation of resources on the one hand, and the noticing and de-homogenising of poor schools and teachers working within those poor schools on the other.

It is clear that the pedagogic, transmitter and acquirer codes require careful *mediation*, in order that schools, teachers and learners and their parents are able to decipher and 'read' the code more penetratingly and hence more productively (see Archer, 2010; Archer et al, 2013 for an example). This is a lesson the provincial authorities might pay attention to. Without this, the potential for tests to be experienced as alien, threatening and coercive remains, and the potential of tests as a

crucial diagnostic device undermined. What is apparent from the analysis of the data is an emergent exercise of pedagogic agency with respect to the tests, particularly in decisions around the distribution of resources. We must note, however, that this pedagogic agency is located only in the ‘recognition’ phase of pedagogic labour, and not yet in the ‘realisation’ phase, that is, that these emergent discriminant recognitions have not in this study been examined to see whether they translate into observable pedagogic practices which can plausibly be said to cause the increase in test results (Bernstein, 1996). The claims made in this paper are therefore only to do with emerging discrimination in recognition judgments.

This paper has thus identified an absence in the literature on tests and testing, an absence that has to do with the *identification of pedagogic judgements and pedagogic agency* that lie between the administration of the test and the test results. In our view, both the critics and the advocates of testing tend to pass over the critical juncture of pedagogic agency, in either, or both, of its recognition and its realisation phases. The advocates simply assume agency, and leave it in the ‘black box’; the critics emphasise disempowerment, which eclipses the possibility for agency. This paper makes a start in identifying the dynamics of this pedagogic agency.

The process of systemic testing in South Africa has a long way to go. Within the strata of underperforming schools in our sample and within a low-stakes testing environment, testing would appear to drive differentiation within and between schools. No large claims are made here for any macro-interruption of the inequity order. Nevertheless, the individualization of teachers and schools, where performance is ‘noticed’, holds the potential to foster greater self-regulation and improved academic outcomes, hence greater micro-equity, for children in these contexts.

References

- Amrein, A.L. & Berliner, D.C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10,18.
- Archer, E. (2010). Bridging the gap: optimising a feedback system for monitoring learner performance. Unpublished doctoral dissertation, University of Pretoria, Pretoria.
- Archer, E., Scherman, V. & Howie, S. (2013). Approaches to effective data use: does one size fit all? In K. Schildkamp, M. K. Lai & L. Earl (eds.) *Data-based decision making in education*, 91 – 112. Dordrecht: Springer.

Au, W. (2008). *Unequal by Design. High Stakes Testing and the Standardization of Inequality*. New York: Routledge, Taylor & Francis.

Baker, E.L., Barton, P., Darling-Hammond, L., Haertel, E., Ladd, H., Linn, R., Ravitch, D., Rothstein, R., Shavelson, R. & Shepard, L. (2010). *Problems with the use of student test scores to evaluate teachers*. EPI Briefing paper 278. Washington, D.C.: Economic Policy Institute.

Barrett, B. D. (2008). No Child Left Behind and the assault on teachers' professional practices and identities *Teaching and Teacher Education*, 25,8, 1018-1025.

Beck, J. & Young, M. (2005). The assault on the professions and the restructuring of academic and professional identities: A Bernsteinian analysis. *British Journal of Sociology of Education*, 26, 2, 183 – 197.

Bernstein, B. (1996). *Pedagogy, Symbolic Control and Identity: Theory, Research, Critique*. London: Taylor & Francis.

Clarke, M., A. Shore, K. Rhoades, L. Abrams, J. Miao, and J. Li. (2003). Perceived Effects of State-mandated Testing Programs on Teaching and Learning: Findings from Interviews with Educators in Low-, Medium-, and High-stakes States. National Board on Educational Testing and Public Policy, Boston College.
<http://www.bc.edu/research/nbetpp/statements/nbr1.pdf>.

Crocker, L. (2004). Teaching for the Test: How and Why Test Preparation is Appropriate. In R. P. Phelps (Ed.) *Defending Standardized Testing*, 159–174. Mahwah, NJ: Lawrence Erlbaum Associates.

de Clercq, F. (2013). Professionalism in South African education: The challenges of developing teacher professional knowledge, practice, identity and voice. *Journal of Education*, 57.

Fleisch, B. (2008). *Primary education in crisis: Why South African school children underachieve in reading and mathematics*. Cape Town: Juta.

Hardy, I. (2014). A logic of appropriation: enacting national testing (NAPLAN) in Australia. *Journal of Education Policy*, 29:1, 1-18

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 1, 81 – 112.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating*

to achievement. London, UK: Routledge.

Hoadley, U. & Ensor, P. (2009). Teachers, social class, professional dispositions and pedagogical practice. *Teachers and Teacher Education*, 25, pp. 876–886.

Hopmann, S.T. (2008). No child, no school, no state left behind: schooling in the age of accountability. *Journal of Curriculum Studies*, 40:4, 417-456

Hursh, D. (2013). Raising the stakes: high-stakes testing and the attack on public education in New York. *Journal of Education Policy* (iFirst Article 7 Jan 2013), 1-15. doi: 10.1080/02680939.2012.758829

Jones, B.J. (2007). The Unintended Outcomes of High-Stakes Testing, *Journal of Applied School Psychology*, 23:2, 65-86

Labaree, D.F. (2012). School syndrome: Understanding the USA's magical belief that schooling can somehow improve society, promote access, and preserve advantage, *Journal of Curriculum Studies*, 44:2, 143-163.

Lam, David, Cally Ardington, and Murray Leibbrandt. (2011). "Schooling as a lottery: Racial differences in school advancement in urban South Africa." *Journal of Development Economics*, 95(2): 121-136.

Lingard, R. (2011). Policy as numbers: a/counting for educational research. *Australian Education Researcher*, 38(4), 355–382.

Lingard, B., Creagh, S., & Vass, G. (2012). Education policy as numbers: data categories and two Australian cases of misrecognition. *Journal of Education Policy*, 27(3), 315-333. doi: <http://dx.doi.org/10.1080/02680939.2011.605476>

Lingard, B. & Sellar, S. (2013). 'Catalyst data': perverse systemic effects of audit and accountability in Australian schooling. *Journal of Education Policy*, 28,5.

Lobascher, S. (2011). "What are the Potential Impacts of High-stakes Testing on Literacy Education in Australia?" *Australian Journal of Language & Literacy*, 34, 2: 9–19.

Madaus, G., M. Russell, and J. Higgins. (2009). *The paradoxes of high stakes testing: how they affect students, their teachers, principals, schools and society*. Charlotte, NC: Information Age Publishing.

Phelps, (2006). "Characteristics of an Effective Student Testing System," *Educational Horizons*, 85, 1.

Phelps (2012). The Effect of Testing on Student Achievement, 1910–2010, *International Journal of Testing*, 12:1, 21-43

Polesel J., Rice, S. & Dulfer, N. (2013). The impact of high-stakes testing on curriculum and pedagogy: a teacher perspective from Australia, *Journal of Education Policy*, DOI: [10.1080/02680939.2013.865082](https://doi.org/10.1080/02680939.2013.865082)

Ravitch, D. (2010). *The death and life of the great American school system*. New York: Basic Books.

Singh, P, Martsin, M., Glasswell, K. (2014). Dilemmatic spaces: High-stakes testing and the possibilities of collaborative knowledge work to generate learning innovations. *Teachers and Teaching: Theory and Practice*.

Sloane, F., and A. Kelly. (2003). Issues in High-Stakes Testing Programs. *Theory into Practice*, 42, 1: 12–17.

Tanner, D. (2013). Race to the top and leave the children behind. *Journal of Curriculum Studies*, 45,1, 4-15.

Santiago, P., G. Donaldson, J. Herman, and C. Shrewbridge. (2011). OECD Reviews of Evaluation and Assessment in Education: Australia. Paris: OECD.

Sapire, I. & Herholdt, R. (2014). The role of error analyses in teacher development: Using the Annual National Assessments as a case in point. Paper presented at Knowledge Week 2014. University of Witwatersrand, Johannesburg.

Spaull, N. (2012). [Poverty & Privilege: Primary School Inequality in South Africa](#) (July 2012 – Working Paper)

Taylor, N. & Taylor, S. (2013). Teacher knowledge and professional habitus. In: N. Taylor et al (Eds.). *Creating effective schools*. Cape Town: Pearson.

Taylor, S., and D. Yu. (2009). The importance of socio-economic status in determining educational achievement in South Africa, Working Papers 01/2009. Stellenbosch University, Department of Economics.

Van der Berg, S., and M. Louw. (2006). Lessons learnt from SACMEQII: South African student performance in regional context. Paper presented at the Investment Choices for Education in Africa Conference, September 19–21, in Johannesburg, South Africa.

Van der Berg, S. & Shepherd, D. (2010). Signalling performance: Continuous assessment and matriculation examination marks in South African schools. Stellenbosch Economic Working Papers 28/2010.

Venkatakrishnan, H. and Spaul, N. (2014). *What do we know about primary teachers' mathematical content knowledge in South Africa? An analysis of SACMEQ 2007*. Stellenbosch Economic Working Papers 13/2014.

World Bank. (2013). Review of the Annual National Assessment (ANA) program, South Africa.

Yates, L. (2013). Revisiting curriculum, the numbers game and the inequality problem. *Journal of Curriculum Studies*, 45,1, 39-51.

Young, M. & Lambert, R. with C. & M. Roberts (2014). *Knowledge and the future school: curriculum & social justice*. London: Bloomsbury.