

National Standards for New Zealand's Primary and Intermediate School Pupils

MARTIN THRUPP

Abstract:

A key education policy plank for the National Party going into the 2008 election is to be National Standards for primary and intermediate pupils in reading, writing and numeracy. These are seen to be more acceptable than the national testing which occurs in England, the United States and other countries. But are they really more acceptable? This article will review evidence about the perverse effects of national testing, consider what is known about the National Standards proposed for New Zealand, and assess whether they are likely to avoid many of the same damaging effects on schools and pupils. It is argued that although the National Party is trying to distinguish its policies from national testing, the distinctions are not yet significant enough to prevent the problems which have been experienced overseas.

This article offers a critical perspective on the National Party's plans to introduce National Standards for primary and intermediate pupils in reading, writing and numeracy, if elected in 2008. The details of National's plans are sketchy at the time of writing, but the party's emerging policy intentions are still worth discussing, because they may indeed represent a very significant adverse development for primary schooling in New Zealand. By the time this article is published, it may have become clear that New Zealand is not going to go down the path of national testing, and in this case the article will serve to remind us why New Zealand does well to avoid it. A more worrying scenario is that some form of national testing *will* come in, and the problems it has raised in other countries will become all too relevant to New Zealand as well.

The first two sections of this article give a brief review of the import of national testing in England and the United States.¹ What becomes clear from reviewing its effects in these settings is just how damaging it is. It is important to recognise at the outset that the damage created by national testing is only partly caused by testing *per se*. It is exacerbated by related target-setting and accountability regimes which make the testing "high stakes". Setting targets against which to monitor performance is a key New Public Management tool (Boston et al., 1996) and an important regulatory technique for "steering from a distance." Under New Labour, target-setting has become a key feature of public policy in England, and schooling has become saturated with required targets at every level. With the passing of the federal No Child Left Behind (NCLB) Act, it is also becoming increasingly important in schooling in the United States as well. Although national testing (and its related target-setting and accountability regimes) may often seem to deliver higher student performance, in fact there are problems with its *reach, sustainability, authenticity and credibility*. The testing policies imply an acceptance of continued student failure, there are weaknesses in the content of the tests and their assessment, and there is evidence of plateauing as the gains due to "teaching to the test" are used up. The pressures of national testing towards narrowing the curriculum and fabrication of results are difficult to avoid. Externally-set targets are generally perceived by practitioners as forced or "inauthentic" and hence often inappropriate. Moreover there is much evidence that those worst affected are students from low socio-economic status (SES) and minority backgrounds, as well as those who have special needs of various kinds.

In the final section of the article I will turn to what is known so far of the National Party's policies on national standards and ask whether these will have similarly adverse consequences if it is elected into power in 2008.

National Testing in England

One problem with national testing in England is how accepting of student failure the targets for national test passes are, especially at Key Stage 4. In England only about 50 percent of students have been expected to get 5 A-C grades (i.e., pass grades) in the General Certificate of Secondary Education (GCSE), rising to 60 percent as a target by 2008. Put another way, it is accepted that some 50 percent (or at best 40 percent), of the student population will fail. As Torrance has commented, "if we assume a reasonable indication of a sound secondary education is 5 GCSE A* - C grades, including English and Maths then

around 50% of the school population is still not achieving this" (2003, p. 922). The scale of this failure and its links to social class and school composition have been highlighted by Ofsted² Chief Inspector David Bell, who has noted that only 32 percent of children whose parents have "routine occupations" are likely to leave school with five good GCSE passes, compared with 81 percent of children from more advantaged homes (Bell, 2003).

A second, related problem is that test gains are clearly proving unsustainable over time, that is, they are "plateauing". This is particularly obvious from the Key Stage 2 results, which rose steadily from 1995-2000 but then plateaued, such that by 2005 the number of students reaching Level 4 in numeracy and literacy had still not reached the targets of 75 percent and 80 percent, respectively, which should have been reached by 2002. A frequent interpretation of this plateauing trend is that it suggests the "using up" of capacity to improve, due to "teaching to the test", rather than improved learning and teaching *per se*. For instance, Tymms (2004) has argued that test data have been used in a very "high stakes" fashion so that the pressure created makes it hard to interpret them. "Teaching test technique must surely have contributed to some of the rise, as must teaching to the test" (p. 492). Similarly, Wiliam (2001) has argued that, "rising test scores demonstrate little more than teachers' increasing abilities to teach to the tests, and to the power of high-stakes tests to distort the curriculum" (p. 14). Certainly, even before looking in detail at relevant school processes, one is led to the conclusion that the pattern of test results supports the teaching to the test thesis. Over the late 1990s when mathematics and English results were rising, results in science, the other core subject, rose at a similar rate. This was despite having no "science hour"; in fact teachers reported they were teaching less science because of the "literacy hour" and "daily maths lesson". So the similar rates of improvement for mathematics, science and English are unlikely to have been because of government strategies, because they did not occur in science. It is much more likely that teachers were getting better at teaching to the tests.

These are the most obvious problems indicated by patterns in the English test results data, but there are others. Tymms (2004) has used results from several different studies and an analysis of the standard-setting procedures to provide a comprehensive analysis of the officially reported gains in achievement at Key Stage 2. His main conclusion is that national testing has failed to accurately monitor standards over time because:

- statistical procedures were faulty, and not corrected until 2000/01;
- the form of the national tests has changed over time;
- the curriculum has inevitably changed, and therefore so has test content.

Furthermore Wiliam (2001) has pointed out that curricula are always changing and that what is *actually taught* in schools changes even if the "official" curriculum has not changed.

All of this should be enough to warn against taking achievement gains, and hence the success of national testing, at face value. This view is reinforced when the school processes underlying the achievement figures are considered. A key concern is that increased emphasis on narrow targets has reduced the curriculum in various ways. To begin with, national testing and target-setting has required a closer specification of formal content and teaching methods from central government. This was most obvious in the core primary subjects of English and mathematics when a National Literacy Strategy (NLS) and a National Numeracy Strategy (NNS) were introduced at Key Stages 1 and 2 (subsequently the NLS has been extended to Key Stage 3 in secondary schools). Secondly, there is much evidence that schools and teachers focus on the test/target. Indeed Gillborn and Youdell (2000) discuss what they call the "A-to-C economy" in which "almost every aspect of [secondary] school life is re-evaluated for its possible contribution to the headline statistic of the proportion of pupils attaining at least 5 higher grade GCSE passes" (p. 12).

The pressure to meet targets in national tests has a variety of unjust and perverse consequences. It encourages schools to recruit bright, middle class, "able" children and avoid taking on "expensive" Special Educational Needs (SEN) and excluded students wherever possible (Ball 2003; Bagley et al., 2001). Once in schools, children are seen to be similarly commodified or marginalised through decisions around "setting" and testing. For instance, Gillborn and Youdell (2000) note the occurrence of "educational triage" where decisions are made to focus on some students at the expense of others, depending on whether or not they are seen to have the potential to enhance their school's A-C position. Pressure to meet targets for national tests may also often lead to "impression management" by fabricating results (Ball, 2001), and contribute to a general decline in the vitality and creativity of teaching (Gewirtz 2002).

Other criticisms revolve around:

- intensification of teacher and management workloads;
- the de-motivating effect of ambitious test targets on the students and schools that cannot reach them; and
- the difficulty of test targets which rise year on year when the characteristics of student year groups can vary markedly.

There are concerns too that “the over-determined New Labour classroom may well produce a generation of young people marred by what Hugh Lauder calls ‘trained incapacity’ ” (cited in Ball, 1999, p. 202). Target-setting and testing may represent a considerable opportunity cost in terms of time, energy and funding which could be better employed in advancing clearer goals of educational and social justice.

National Testing in the United States

The United States has been moving towards a national target-setting and testing regime which has numerous similarities to that in England. Until 2002, states varied in the extent to which they were willing to implement standards and standardized testing, and in the significance given to the tests. In 2002, the No Child Left Behind (NCLB) Act mandated that *all* states conduct standardized tests in reading and mathematics in grades three through eight, and by 2008, add science and extend the testing to grades nine through twelve. Schools now face severe penalties for failing to increase test scores sufficiently, including potentially being converted into charter (privately-administered, publicly-funded) schools. Most of the Act’s requirements immediately went into effect, including requiring all states to develop standardized tests and assessment systems in order to determine whether schools are making “adequate yearly academic progress” (AYP).

Under NCLB, schools must not only develop and assess students; they must make public the aggregated test scores for groups of students delineated by gender, race and ethnic group, and with or without disability. Because schools may have several ethnic and racial groups, all of whom must be represented by gender and ability status, schools typically have to generate and report the scores for dozens of subgroups. The test scores for each group are then compared to the state’s testing requirements to determine whether the group is making Adequate Yearly Progress (AYP). If any one group fails to make AYP, the

entire school is designated as failing. Sometimes, shifting a single child’s score from one group to another (for example, by changing a student’s disability status) can mean the difference between a school failing or passing. Schools that fail to make AYP face significant sanctions which become more severe year by year unless there is an improvement (U.S. Department of Education, 2002a). If schools do not make AYP for two consecutive years, they must be identified as “in need of improvement” and students allowed to transfer to another public school with transportation provided by the failing school. By the fourth year, “failing schools” must take corrective action that includes one of the following: replacing the school staff, implementing a new curriculum, “decreasing management authority at the school level, appointing an outside expert to advise the school, extending the school day or year, or reorganizing the school internally” (U.S. Department of Education, 2002b, p. 6). Further, if a school fails to make progress for five years, the school district must initiate plans for fundamental restructuring:

This restructuring may include reopening the school as a *charter school*, replacing all or most of the school staff who are relevant to the failure to make adequate progress, or turning over the operations either to the state or to a *private company* with a demonstrated record of effectiveness. (U.S. Department of Education, 2002b, p. 7, italics added).³

The consequences of testing and target-setting in the United States have many parallels to the situation evident in England. One extensive review of anecdotal evidence from all states concludes “that the over-reliance on high-stakes testing has serious negative repercussions that are present at every level of the public school system” (Nicols & Berlinger, 2005, p. i).

In New York, almost every recent standardized examination has been criticized for having poorly constructed, misleading, or erroneous questions, or for using a grading scale that either over- or under-states students’ learning. Critics argue that the degree of difficulty of an examination has varied depending on whether the State Education Department (SED) has wanted to increase the graduation rate (and therefore makes the exam easier) or appear rigorous and tough (and therefore makes the exam more difficult). The passing rate for the exam can be increased or decreased simply by adjusting the cut score, turning a low percentage of correct answers into a pass or a high percentage of

correct answers into a failure. The SED has also been criticized for the way in which it constructs the test questions. For example, a recent English examination received national censure for removing from literary passages references “to race, religion, ethnicity, sex, nudity, alcohol, even the mildest profanity and just about anything that might offend someone for some reason” (Kleinfield, 2002, p. 1; see Hursh, 2005, for further details).

However, even if the tests were well constructed and valid, the yardstick by which schools are measured – adequate yearly academic progress (AYP) – often discriminates against schools serving students of colour and living in poverty. The determination of whether a school is making AYP tells us little about whether a school is improving. Not only can the validity of the tests be questioned, but the determination of success or failure may have little to do with whether the school is improving. Under NCLB, *every state*, with the approval of the federal Department of Education, determines, for *every test*, what knowledge and skills students need to demonstrate proficiency. States can, therefore, make achieving proficiency more or less difficult. However, for all states and every school, all students (regardless of ability or proficiency in the English language) are to achieve proficiency by the year 2014.

Contrary to a common sense interpretation of AYP, schools are not evaluated on whether their test scores are improving, but whether their aggregated and disaggregated test scores exceed a minimum yearly threshold that will gradually increase over the next decade. Consequently, a school is considered to be passing as long as its scores exceed the threshold, even if the scores fall. Similarly, schools that begin with initially low test scores may be considered failing even if they significantly improve their test scores, as long as those scores remain below the threshold. Thus, achieving AYP may have nothing to do with whether a school’s test scores rise or fall; achieving AYP depends only on exceeding the minimum threshold.

Because test scores strongly correlate with a student’s family income, a school average score is more likely to reflect its students’ average family income, rather than the quality of teaching or the nature of the curriculum. Consequently, most failing schools in New York are found in poor, urban school districts. Almost all (83%) of the failing schools are located in the big five urban districts: New York City, Buffalo, Rochester, Syracuse, and Yonkers (New York State School Boards Association, 2002). Most of the remaining failing schools are in smaller urban

districts. The failure rate of schools in the urban districts is high, particularly at the middle school level. In Rochester, for example, all the middle schools failed. To NCLB’s testing requirement that schools demonstrate improvement for all disaggregated groups of students on all the tests, Florida added the further stipulation that no school that had been assigned by the state a grade of a D or F (on the annual rating of A through F) could meet AYP. Not surprisingly, 90 percent of Florida’s public schools were designated as failing to meet AYP, and 100 percent of districts failed (Pinzur 2003).

Because of the pressure to raise test scores, particularly in the urban school districts, teachers are compelled to teach the skills and knowledge that will be tested, neglecting other usually more complex aspects of the subject, and indeed some subjects altogether. In an ethnographic study of schools in Chicago, Lipman (2004) documents how the testing requirements undermined the critical literacy goals of a bilingual school and frustrated creative, dedicated teachers. She describes how an elementary school, with a student population of over 90 percent Mexican-American, had to shift its focus away from using the students’ own culture to develop critical literacy, and focus instead on test preparation. One teacher stated that she devoted the first half of the school year to encourage her students to write intensively and read sophisticated literature but then, for the third quarter, shifted to test preparation. Test preparation included getting students “used to the format of a short, mediocre selection of writing ... to get them to recognize this type of question is asking you for some really basic information you can go back to look for” (Lipman, 2004, pp. 110-111). The teachers, Lipman writes, experience “the contradictions and conflicts ... between their efforts to help students see knowledge as a tool to analyze the world, and the process and practice of preparing for standardized tests” (p. 111).

In Chicago, as in England, administrators instructed teachers to put their efforts into raising the test scores of those students who are closest to passing the standardized tests (Gillborn & Youdell, 2000). As one teacher said:

They tell us ... “We don’t want high kids and we don’t want the lowest kids, we want the kids that are just about to pass the IOWA [standardized] test.” So here you have a third or a fourth of your classroom really needs help to be ready for that next grade level and they don’t get to go. (Lipman, 2004, p. 82)

Such educational “tragedy” (Gillborn & Youdell, 2000) exacerbates educational inequality, as the students who either pass or are close to passing the test become valued commodities, and those students who need the most help are left to fend for themselves.

McNeil (2000) illustrates how the emphasis on tests and test scores undermined exemplary schools and teachers in Houston, Texas. She documented how after the Texas Assessment of Academic Skills (TAAS) was implemented, previously successful schools began to expect less of their students as they prepared them to pass the more basic skills required on the tests. Rather than, for example, teaching students to write well, teachers taught students to write the five-paragraph essay with five sentences in each paragraph that would receive passing grades on the standardized tests. Because culturally advantaged middle- and upper-class students are likely to rely on their cultural capital to pass the exams, it is disadvantaged students who receive the additional drilling. Unfortunately, learning to write five-sentence five-paragraph essays does not transfer well to literacy required beyond the test and outside of school. By expecting less of disadvantaged students, they induce them to fall further beyond.

Lowered expectations are not the only problem. Schools emphasizing test preparation are likely to devote most of their curriculum budget to test-prep materials rather than the enriched resources students need. Further, schools, in focusing on test preparation, are likely to reduce or eliminate subjects that are not being tested, including the arts and sciences. In Texas, because science is not tested in the early grades, it is no longer taught.

Rather than ensuring that more students do well, the pressure to raise test scores also encourages schools to force weak students out of school before they take the required exam. In Texas, urban students are more likely to be retained in school, especially in ninth grade, the year before the required TAAS exam is first given. Students who are repeatedly retained are likely to give up and drop out of school. Haney (2000), in his study of the Texas education reforms, concludes that for the year 1996-97, 17.8% of students were being retained in ninth grade (24.2% of African-American and 25.9% of Hispanic students) and that only 57.6% of African-American and 52.1% of Hispanic ninth-grade students were in twelfth grade four years later (Part 5, pp. 8-9).

As noted earlier, schools in Texas have faced a double-edge sword – they need to raise test scores, but face possible sanctions for high dropout rates. Rodney Paige, superintendent of the Houston

Independent School District (HISD) since 1994, found a “fabricating solution” to this problem. He ordered principals not to list a student as dropping out but as having left for another school or some other reason other than dropping out. Such creative book-keeping resulted in the district claiming a greatly reduced dropout rate and winning awards for excellence (Winerip, 2003). Eventually critics claimed that the dropout rate was covered up, and research has indeed revealed the rate to be much higher than the official one. Significantly, almost all of the students that were being pushed out were at-risk students and minorities. A state investigation into 16 high schools revealed that of 5,000 students who left school, 2,999 students should have been reported as dropouts and were not (Winerip, 2003). States other than Texas have also responded to the pressure to raise test scores by surreptitiously pushing students out of school. In New York, students are being pushed out of schools to raise test scores and then, rather than being counted as dropouts, they are listed as having transferred to an alternative school or working on a Graduate Equivalency Diploma (GED), a diploma achieved not by attending school but by passing an exam (Lewin & Medina, p. 1).

We might expect that when the curriculum is narrowed and simplified, students who score low on tests are abandoned, poorly constructed tests lead to mass failures, and students are pushed out of schools, then the achievement gap will grow rather than reduce. Evidence from New York supports this conclusion. From 1998 to 2000, the number of students dropping out increased by 17 percent. A recent report for the Harvard Centre for Civil Rights concluded that New York now has the lowest graduation rate of any state for African-American (35%) and Latino/a (31%) students (Orfield, Losen, Wald & Swanson, 2004). In New York City only 38 percent of all students graduate on time, making it the fifth worst of the 100 largest cities in the nation (Winter, 2004). According to another recent study, New York’s graduation rate ranks 45th in the nation (Haney, 2003). The tests have also negatively affected English language learners, who moved from the highest diploma-earning minority in 1996 to the highest dropout minority in 2002 (Monk, Sipple & Killen, 2002). Lastly, dropouts among students with disabilities have increased from 7,200 in 1996 to 9,200 in 2001.

The quantitative evidence from Texas is more contradictory and contested. The state reports that the mean student test score and percentage passing the TAAS exam has increased, the differences between the mean test scores for White, African-American and Hispanic

students have decreased, and school drop-out rates have declined. However, Haney (2000) investigated the Texas data and revealed how the higher test scores were achieved. First, while students who are in special education must take the TAAS, their scores are not included in those reported by the school. Therefore, if students whose scores might negatively affect the overall school score can be excluded by placing the student into special education, we might expect, after TAAS was implemented, that the percentage of students in special education to increase. Haney (2000) shows that for the first four years in which TAAS was implemented, the percentage of special education students increased from 4.5% to 7.1%. This indeed seemed to happen.

A second way to increase test scores has been to retain students in grades prior to tenth grade, the grade in which students first take the TAAS, providing them another year to prepare for the test. Haney's data reveal that the retention rate for previous grades has increased significantly, particularly in the ninth grade. In 1996-97, 25.9% of Hispanic, 24.2% of African-American and 17.8% of White students were retained in ninth grade. Of course, grade retention also increases the likelihood that a student will drop out of school. Rather than relying on the dropout rate reported by schools and school districts, Haney (2000) compared the percentage of students in ninth grade with the percentage of students in twelfth grade four years later. His data reveal that in fact there has been a significant increase in the dropout rate in Texas. Therefore, Haney concludes, the Texas "miracle" was really the Texas "mirage". Test scores have increased because students are increasingly likely to be retained in previous grades, or have become so discouraged that they have quit school altogether. In addition, extra students have been placed in special education so that their lower scores would not be included in the reported scores. In sum, Texas schools have raised test scores by retaining students or removing them from the pool of test takers. Rather than increasing education achievement, the whole exercise has meant that fewer students have the opportunity to receive an education.

Even as schools have manipulated the scores by limiting who takes the examinations, the higher average score may only mean that the students are performing better on the tests, not that they are learning more. While students' scores on the TAAS exam have been increasing, their scores on nationally administered tests, such as the university admissions examinations, have been decreasing. Researchers investigating the phenomenon explained:

The discrepancy in performance has a lot to do with the differences in the tests. TAAS was designed to make sure that students learned at least the basics of the state curriculum. The [university admissions] tests, on the other hand, assess students on advanced academic skills needed for college. (Markley, 2004)

Developments in New Zealand

New Zealand's recent history of target-setting would have to be considered "soft-touch" compared to that of England and the United States. The Education Act (2001) required New Zealand schools to describe their priorities for improvement and report on progress against targets. Schools also have to report on progress against their targets in the analysis of variance section of their annual reports. But schools set their own targets, rather than have targets set for them as in England and the United States. There is also a considerable amount of freedom in the way in which targets are set, and there are not yet punitive consequences for missing such targets.

Despite the limited development of target-setting in New Zealand, the National Party has a history of favouring more national testing in schools. Over the last decade, plans for national testing were advanced by National over 1997, dropped in March 1998, and picked up again in September 1999, when National announced a piloting of national testing in literacy and numeracy for 9- and 11-year olds. This was just prior to the 1999 general election, but in the event the National Party lost the election and the pilot was quickly scrapped when Labour got in. National again raised the prospect of national testing just before the 2005 election when Bill English MP announced a School Reporting policy which was intended to give parents more access to information about how their child was performing at school. It was noted that schools currently use a range of tests, and many schools set targets and report against them, and the argument was advanced that without developing any new tests parents and students could benefit from this practice. National went on to say it would:

- Set national standards in literacy and numeracy, using agreed benchmarks in existing tests;
- Require schools to use Plain English Reporting in literacy and numeracy up to Year 10 (4th Form) level. These Plain English Reports will show what progress students have made relative to national standards;

- Require schools to report progress of their school towards national literacy and numeracy standards, including comparisons with similar schools; and
- Require the Ministry of Education to provide parents with direct access to the 20 indicators of school performance already collected and available to schools through the “School Smart” system.

These plans were not advanced because the National Party lost the 2005 election, but they have resurfaced in the party’s latest policies. In a speech given in April 2007, National’s leader John Key announced that national standards would be the cornerstone of [National’s] education policy. The rationale of the proposal is that too many children are failing, and that without clear standards and reporting, expectations are too low:

[The Education Review Office (2005)] said that the area where the school system is least effective is in identifying these students. That finding is just staggering – the school system is not identifying these students. Now, you would have thought that was happening already: that schools were testing kids, seeing whether they had developed the skills appropriate to their age or year level, and setting off alarm bells if they weren’t. But obviously a lot of schools are not doing this.

I don’t think a child’s home circumstances – be they from a poor family, a single-parent family or any other sort of family – should determine our expectations of how they will do at school. The Māori Party talk about a “deficit mentality” that permeates our education system. This is an expectation that some kids, particularly Māori kids, will fail. I don’t want that mentality to continue. I want us to be aspirational. I have high expectations for every child and I expect schools and teachers and students to share those expectations. (Key, 2007)

National’s policy response is to propose three key requirements for all primary and intermediate schools:

- *Clear National Standards:* Set national standards in reading, writing and numeracy. The standards will describe all the things children should be able to do by a particular age or year at school. They will be defined by benchmarks in a range of tests.
- *Effective Assessment:* Require primary schools to use assessment programmes that compare the progress of their students with other

students across the country. Schools will choose from a range of tests, but there won’t be national exams.

- *Upfront Reporting:* Give parents the right to see all assessment information, and to get regular reports about their child’s progress towards national standards. Schools will also have to report each year on the whole school’s performance against national standards. (“National Standards cornerstone of education policy”, 2007)

Further details in a Q & A page on the National Party website (National Party, 2007) suggest national standards will essentially be “floor” targets:

How do National Standards differ from the National Curriculum?

National standards will focus teachers on the bottom-line skills that every child should be mastering in reading, writing and mathematics. The standards will be explicitly defined by benchmarks in tests. The Curriculum is much broader.

There are details about the tests to be used:

How can you enforce National Standards across the country if you don’t have one national test?

All schools will be required to choose tests that have been benchmarked against national standards. Teachers will have clear guidelines about what children need to achieve in a test to reach the national standard. They will be required to use tests to work out whether their pupils have reached the national standard.

The National Party is keen to differentiate its plans from national testing:

How is this different from national testing?

We won’t require schools to use one government-approved test. Instead schools will choose from a range of tests and integrate those tests into their teaching.

Do any other countries have systems like this?

It is agreed all over the world that quality assessment underpins quality teaching and quality learning. It’s also acknowledged that teachers who set clear achievement targets for their students get good results. Countries like the UK and the US have gone as far as to set universal national tests; we don’t think that’s necessary in New Zealand.

Discussion and Conclusion

Evidence on the impact of national testing in England and the United States suggests it makes education less equitable and less authentic. National is clearly keen to differentiate its proposals for National Standards from national testing, but it needs to be asked whether the differences will be enough to prevent its National Standards being much different from, or any less problematic than, the national testing approaches used overseas.

I would suggest that National is probably putting too much store in allowing for a *range* of tests as differentiating its approach from national testing. In practice schools will probably fall back onto three tests. First, the *Progressive Achievement Tests* (PATs), which New Zealand teachers have long used to assess reading, mathematics, and listening comprehension. Secondly, the *Supplementary Tests of Achievement in Reading* (STAR), which provide additional assessment in areas such as word recognition, sentence comprehension and vocabulary. Thirdly, and perhaps most significantly, *Assessment Tools for Teaching and Learning* (asTTle). This is a computer-based assessment system made available free of charge to all New Zealand schools by the Ministry of Education. It is used for assessing literacy and numeracy in both English and Māori, and provides information about a student's level of achievement, relative to the curriculum achievement outcomes, for Levels 2 to 6, and gives national norms of performance for students in Years 4 to 12. It is clear that the Ministry of Education is investing heavily in asTTle. In 2009 it will launch e-asTTle, which is web-based and:

has made advances in standard-setting methodologies. It gives teachers a realistic picture of how well each student, class or school is doing compared to the national average and the curriculum requirements (including curriculum levels). It allows comparisons to other groups such as gender, ethnicity, English as a second language, or "schools like mine". ("Why is e-asTTle ground-breaking?", 2007)

The strengths of e-asTTle need to be weighed against the concern that it also clearly provides the kinds of assessment technology which will make it much easier for a future government to collect assessment data across all schools, publish league tables and set up the kinds of performative pressures and perverse consequences evidenced overseas, but which New Zealand primary schools have been relatively free from thus far. In the past, these tests have all been used at the discretion of

the classroom teacher, to enhance their teaching and provide appropriate feedback to their parents. National and regional norms are provided for comparative purposes, but the onus is still on the teacher, not the system.

Another feature of the proposed National Standards policy is that it currently stresses minimum or "floor" targets: "the standards will describe all the things children should be able to do by a particular age or year at school." Punitive floor targets are an important part of national testing in England and even more so under NCLB in the United States. It might be, however, that the policy will evolve into more "aspirational" individual targets for all pupils. Whatever the nature of the targets, a key issue will be whether they will be set by schools or externally determined. Targets set by schools might be regarded as lacking rigour, but they will allow schools more room to respond to their own context and help to avoid the fabrication which schools and teacher engage in when faced by inappropriate but high stakes targets.

A related development is that the incumbent Labour-led government is introducing the first of a series of curricula progressions that can be used to tell teachers what students should know and be able to do at particular ages. ("Progressions join literacy toolkit", 2007; Ministry of Education, 2007). Although the progressions are presented as an inherent part of wider curriculum reform, they are also likely to be a way of responding to National's claims that there are not enough "standards" now. Indeed, if the issue of standards in schools is seen to have electoral appeal, the danger must be that political manoeuvring between the major parties will lead to testing and target setting becoming ratcheted up over time beyond what either of the major parties would currently consider desirable.

Another concern with the proposed shift to national standards is that the many advantages of national monitoring will become overlooked. Every year since 1995 (when it began under a National Government), the National Education Monitoring Project (NEMP) based at the University of Otago has been reporting on what New Zealand primary school children know and can do, covering all areas of the curriculum over consecutive four-year assessment cycles. New Zealand has been well served by NEMP's ability to work productively with teachers to monitor achievement over time and provide assessment information of many kinds, including scores on creative performance tasks, and not just the traditional paper-and-pencil tests. At the same

time, the ability of this project to satisfy the needs of policymakers for monitoring achievement has to be seen as one of the key reasons New Zealand has so far been able to keep national testing at bay, along with all the perverse effects already noted.

At the time of writing it remains to be seen quite where the National Party will go with its National Standards, or even if it will lead the next government. Nevertheless this article has argued that New Zealand would do well to continue to avoid national testing of the type implemented in many jurisdictions overseas. National's efforts to distinguish its policies from national testing are laudable, but it remains to be seen whether the distinctions are significant enough to prevent the problems experienced elsewhere. Further detail, when it becomes available, should illuminate just what National Standards would involve. It is to be hoped that New Zealand will continue to take a more progressive approach to monitoring student progress, such as it already has with the NEMP programme, and bypass the overseas national testing regimes discussed in this article.

Notes

1. This part of the paper is based on Thrupp, M., & Hursh, D. (2006). The limits of managerialist school reform: The case of target-setting in England and the USA. In H. Lauder, P. Brown, J-A. Dillabough, & A. H. Halsey (Eds.), *Education, Globalization, and Social Change*. Oxford: Oxford University Press. The section on "The costs of national testing in the US" was written by David Hursh, and I am grateful for his permission to use material from it here. Hursh expands on his critique of testing in the US in his new book: Hursh, D. (2008). *High-stakes testing and the decline of teaching and learning: The real crisis in education*. Lanham, MD: Rowman and Littlefield.
2. At the time when Bell was Chief Inspector, Ofsted stood for the Office for Standards in Education. In 2007, Ofsted's remit was widened from schools and early childhood centres to include children's services and employment-based training, and it became the Office for Standards in Education, Children's Services and Skills.
3. There are similar levels of sanctions for failing school districts (See U.S. Department of Education, 2002b, pp. 6-7).

References

- Ball, S. J. (1999). Labour, learning and the economy: A "policy sociology" perspective. *Cambridge Journal of Education*, 29(2), 195-206.
- Ball, S. J. (2001). Performativities and fabrications in the education economy: Towards the performative society. In D. Gleeson & C. Husbands (Eds.), *The performing school*. London: Routledge Falmer.
- Bell, D. (2003, November 20). *Inequality and education: Must urban schools fail?* Anniversary lecture of Ofsted. Report entitled *Access and Achievement in Urban Education*.
- Boston, J., Martin, J., Pallot, J., & Walsh, P. (1996). *Public management: The New Zealand model*. Auckland: Oxford University Press.
- Education Review Office. (2005). *Annual report for the year ending 30 June 2005*. Wellington: ERO.
- Gewirtz, S. (2002). *The managerial school*. London: Routledge.
- Gillborn, D., & Youdell, D. (2000). *Rationing education: Policy, practice, reform, and equity*. Buckingham: Open University Press.
- Haney, W. (2000). The myth of the Texas miracle in education, *Education Analysis Policy Archives*, 8(41). Accessed April 20, 2005 from: <<http://epaa.asu.edu/epaa/v9n2.html>>.
- Haney, W. (2003, September 23). *Attrition of students from New York schools*. Invited testimony at public hearing "Regents learning standards and high school graduation requirements" before the New York Senate Standing Committee on Education, Senate Hearing Room, New York, NY.
- Hursh, D. (2005). The growth of high-stakes testing in the US: Accountability, markets and the decline in educational equality, *British Educational Research Journal*, 31(5), 605-622.
- Key, J. (2007, April) *Encouraging Success: Confronting Failure*. Speech to Auckland College of Education. Accessed February 29, 2008, from National party website: <<http://national.org.nz/Article.aspx?ArticleID=9817>>.
- Kleinfield, N. (2002, July 31). The elderly man and the sea? Test sanitizes literacy texts. *New York Times*, p. 1.
- Lewin, T., & Medina, J. (2003, July 31). To cut failure schools shed students. *New York Times*, p. 1.
- Lipman, P. (2004). *High-stakes education: Inequality, globalization, and urban school reform*. New York: Routledge Falmer.

- Markley, M. (2004, April 1). TAAS scores rose as SAT's fell. *Houston Chronicle*.
- McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- Ministry of Education. (2007). *Literacy learning progressions – draft for consultation*. Wellington: Ministry of Education.
- Monk, D., Sipple, J. W., & Killeen, K. (2001). *Adoption and adaptation: New York state school districts' responses to state imposed high school graduation requirements: An eight-year retrospective*. Education Finance Research Consortium. Accessed March 3, 2002, from: <www.albany.edu/edfin/CR01_MSk_Report.pdf>.
- National Party. (2007). *National Standards Policy Q&A*. Accessed February 29, 2008, from National Party website: <http://national.org.nz/files/_0_National_Standards_Policy_QA.pdf>.
- National Standards cornerstone of education policy.(2007). Accessed February 29, 2008, from National Party website: <<http://national.org.nz/Article.aspx?ArticleID=9818>>
- New York State School Boards Association. (2002). *Title I accountability status updated for March 10, 2003*. Accessed August 28, 2003, from: <www.nyssba.org/adnews/misc/thenewaccountability-5.htm>.
- Nichols, S., & Berliner, D. (2005). *The inevitable corruption of indicators and educators through high-stakes testing*. Tempe, AZ: Arizona State University, Education Policy Studies Laboratory. Accessed from <<http://edpolicylab.org>>.
- Orfield, G., Losen, D., Wald, J., & Swanson, C. (2004). *Losing our future: How minority youth are being left behind by the graduation rate crisis*. Cambridge, MA: The Civil Rights Project at Harvard University.
- Pinzur, M. (2003, August 8). State schools fail to meet new federal test standards: federal, state results differ. *Miami Herald*.
- Progressions join literacy toolkit. (2007, December). *New Zealand Education Gazette* 86(21).
- Torrance, H. (2003). Assessment of the National Curriculum in England. In T. Kellaghan & D. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 905-928). Dordrecht: Kluwer.
- Tymms, P. (2004). Are standards rising in English primary schools? *British Educational Research Journal* 30(4), 477-494.

- U.S. Department of Education, Office of Elementary and Secondary Education. (2002a). *No Child Left Behind: A desk reference*. Washington, DC: U.S. Department of Education.
- U.S. Department of Education, Office of the Secretary. (2002b). *What to know and where to go: A parents' guide to No Child Left Behind*. Washington, DC: U.S. Department of Education.
- Why is e-asTTle ground-breaking? (2007). Accessed February 29, 2008 from Ministry of Education TKI website: <www.tki.org.nz/r/asattle/e-asattle/faq_e.php>.
- Wiliam, D. (2001). *Level best? Levels of attainment in national curriculum assessment*. London: The Association of Teachers and Lecturers.
- Winerip, M. (2003, August 13) The “zero dropout” miracle: Alas! Alack! A Texas tall tale. *New York Times*, p. B7.
- Winter, G. (2004, February 26) Worst rates of graduation are in New York. *New York Times*.

The author

Martin Thrupp is Professor of Education at the University of Waikato. He spent 2000-05 working in London where he undertook research on the effects of target-setting.