**Horizons and whirlpools: The well travelled pathway of national standards**

**John Hattie, Visible Learning Lab**

**School of Teaching, Learning and Development**

**University of Auckland**

New Zealand is sitting on a cusp in relation to education and schooling. According to the international studies (TiMMs, PIRLS and PISA), we sit in the company of the best in the world – we have one of the most effective education systems in the world. We have a cohort of professional teachers who have significantly contributed to this success, and have a society that still values and is prepared to pay for one of the best government school systems in the world. Sometimes we forget that New Zealand is among the exceptions when compared to many developed countries. The New Zealand school success story did not happen by default, and unless nurtured and esteemed, it will quickly slip. So which way to move forward?

It is definitely not a step forward to adopt policies that are known to be flawed when implemented elsewhere - so why have we recently introduced a policy that is failing in other countries? National standards are often the catch cry of countries that have systems in trouble and see a need to bolster their systems. National standards are typically introduced in the midst of despair, aimed to drive mediocrity out of the system, and the push for national standards is often driven by business and civics elites fearful of losing economic benefits. The aim is to make those responsible for providing education (i.e., schools, teachers) more accountable for their results, and to improve the educational health of the nation. The road to national standards is a road well travelled. Many countries, including the United States, the United Kingdom, and Australia, have walked this path, and the path has ended in limited evidence of any improvement in students' educational achievement, difficulties in removing the national standards policy when it is shown that it has no or a negative effect on student achievement and perverse effects in schools. The question is whether New Zealand intends to follow this well travelled pathway.

After a tide of negative reports about schooling in the United States, that country looked to national standards to dramatically improve its system. The first President Bush introduced Goals Education 2000, which was picked up by his successor President Clinton. But despite much haggling and blue ribbon committees, the policy was gaining little, if any, traction, and the second President Bush, realizing that these national standards would merely be words with no effect, proclaimed, on his third day in office, that standards without measurement were not worth the words – hence the policy "No Child Left Behind" (NCLB), in which individual states set their own standards and measures. Now, President Obama says he has the answer to the current malaise about NCLB – he announced recently that the problem with NCLB is that there are now 50 standards and 50 tests, and his solution is to have one set of national standards and one national test. There will more of the same effects of testing to the standards but this time only one chance to get the standards correct and the assessments perfectly aligned. Given the consequential effects of such a one-size-fits-all NS and testing policy it too is likely to head down the same paths of failure -- it may pick up the bottom but also will lead to less positive effects on the brightest and thus unlikely to change the mean, while also carrying with it the many other negative effects.

The road from crisis, to pronouncement of lofty goals, to measuring the lack of success in attaining these goals is thus well signposted. The discovery that if you scratch a standard you will find a test soon becomes evident after the heady days of creating the standards. The dilemma, then, of realizing that the national standards and their associated tests may improve the worst in the system, distort the best, but barely change the average score of the nation is depressing for all – especially politicians who backed national standards as the cure-all. Then more drastic measures (e.g., overly prescriptive summative testing, narrowing the curricula, banning social promotion, removing principals) are introduced usually in the name of making sure those who were not performing or correctly implementing the national standards policy (school leaders, principals, teachers) are held accountable.

When we walk a path, we can often see the destination on the horizon. But, on our national standards pathway, *which* horizon are we seeing? Is it the Near Horizon, where the aim is to ensure that schools enhance their teaching and learning to meet the aims of the national standards, or the Far Horizon, where the aim is to ensure that schools meet the aims of the national standards to thence enhance their teaching and learning? Michael Fullan (2003) has used these notions of the Near and Far Horizon to explore many educational reforms, and he argued that the Near Horizon refers to the policy imperatives and the Far Horizon to transformational effects where we see collective informed professional judgment flourishing. If we aim for the Near Horizon (changing our teaching and learning to successfully implement national standards), then my prediction is that the "path well trodden" will be followed with minor deviations. If, however, we aim for the Far Horizon (using national standards to enhance the quality of teaching and learning across the curriculum) then it is likely that we may reduce some of the variance in our system and may even enhance our teaching and learning. In the words of Robert Frost "Two roads diverged in a yellow wood, and sorry I could not travel both." We need to take the road less travelled.

The issue is not the introduction of standards – we have had them for many years. The pre-2009 curriculum was reasonably clear about standards. A feature of that curriculum was the relation between curriculum levels and year of schooling, with the emphasis (appropriately) on the levels – what they meant, how they related to expectations, and the specification of competencies at each level of the curriculum. The new curriculum is more open and less specific about standards at each level – so given the tension between openness of interpretation and more prescriptive national standards, one might need to go (i.e., the national standards or the new curriculum). Openness and prescription do not readily march together.

**The steps in the well travelled path to national standards**

The start of the national standards path is the setting of the standards. At the beginning the standards have been set by committees – and having worked in the standards movement for two decades, I have yet to meet a group that has not wanted to, or enjoyed, setting standards – especially for others. There are key phrases that surround these standards – they must be aspirational, realizable (with hard work and dedication), and what any reasonably country could expect as outcomes for all students. One of the first tasks that I was given when working for the National Board for Professional Standards, which was deciding on standards for highly accomplished teachers, was to review the standards for teachers in all 50 states in the United States. A room full of boxes later, the message was that they were lovely but useless – lovely in adjectives and aspirations and useless in effect. The mantra is "standards without measurement are useless". A further problem is that standards-setting committees rarely use evidence – claiming often that this will come later. Standards not based on evidence are perverse as they may

have little relation to what students actually can do, they may have little relation to a hierarchy or development path, and they have no valid evidence to defend them.

The next step in the path is the implementation of these committee-developed national standards. A fundamental truism in highly developed school systems is that teachers want to do a good job for their students. This is well rehearsed pathway and in New Zealand the reaction of teachers to NCEA is a good example – despite the speed and absurdity of many of the early NCEA policies, teachers just wanted to do the best for their students – hence the unbelievable amount of work they invested, the hundreds of home-grown solutions, and the thousand flowers that bloomed. It was later that there was a realization that the NCEA was not only a change in how we can teach and learn, but was also a solution to an older bankrupt form of assessment (no longer do universities and workplace depend on the single annual examination of performance, it is done more often, formatively as well as summatively, and is based on evidence of performance). This has led to fixing up the measurement problems and the building of a better system. There is no doubt that 2500 schools will *do their best* to implement national standards. There will be a thousand flowers, and the same reaction will follow – the interpretation of the standards will differ across schools. Where are the moderation processes that are essential to implementing standards, and why are they not part of the introduction and not a discovery to be tacked on later? The damage will be done by not seeing the assessment implications of national standards up front. Standards without assessment solutions are ineffective.

When we step into large scale national standards and then see the results, there is another truism that will get in the way of the mighty magic promised by national standards. If you have a sufficiently large sample and plot reading, writing or mathematics achievement, it inevitably looks like a normal distribution. This means that there will always be a group which are among the most difficult to move above any “aspirational” national target. Given the normal curve, we may be able to get 70% of the students to some reasonable standard, but it is a much harder task to get 80% above. In the United Kingdom, the target for its school standards was set at 80% and when it was able to move the attainment to 74%, the policy was deemed a failure – the United Kingdom could not get to its target of 80%. If only policy developers had realized the implications of a normal distribution, it may have succeeded! It is just not reasonable to expect 100% success, so where is the “target” of acceptability, or better, where is the target of “progress”. There has been NO reference to the current status of levels and progress by years (as outlined in the national standards booklets or releases), so again it is a case of standards without reference to evidence and, even more important, standards without any criteria to show where we are and where we would like to be. The chances of successful implementation diminish in the absence of this information. Dogma and ideology can only prosper in such circumstances.

Another common step in the national standards pathway is to find ways to maximize the number of students that reach the standards. All very sensible – but some are more devious than others as schools and the system finds that it is not easy to get all, or even nearly all, above the national standards! Let the games begin!

- The most invidious “game” is to blame the students. “I have taught them, but they have not learned”, “They come from poor homes so I cannot be responsible”, “There is no tail of underachievement, there is a tail of poverty”, “I have Maori/Pasifika/ESL (or many other groups) and they underachieve,” “He’s ADHD/Aspergers/Dyslexic (and so on), so I cannot be held responsible,” “I cannot be asked to set targets, as if the kids don’t reach them it is because they would not put in the effort, do the homework.”

- Then we have “accommodation”. What unseemly dastardly deed could be done in the name of this beautiful word? In the NS context, it means, in fact, excluding students for some reason – these can include special education students, recent immigrants, distressed students in pressure situations, students needing writers or assistants. In the United States, the percentage “accommodated” rose to 27% of all students. The obvious problem accrues – these students become absent in the high stakes tests, but they also lose in the teaching stakes as attention is given to those who will take the tests and thence affect the reputation of the school!
- Another game to be played is retention. If you take a struggling Year 5 student and promote him to Year 6, this student will contribute to lowering the school mean; but if you hold the student back a year, he may do less damage to the Year 5 mean! An analysis of 205 studies on retention can show retention as one of the more systematically negative policies in schooling, and on top of this it is sexist (80% plus of retained students are boys) and racist (80% plus of retained students are students of colour). Schools may think they are gaining a one-off boost to their mean scores at the expense of dramatically increasing the probability of damaging students.
- There are also games about “magic bullets” – if only you try this method or use that resource, it will provide the solution – which may be fine if the bullet is evidence based, but watch for the mandating of this bullet (e.g., literacy hour, minimum hours per day on literacy and numeracy, phonics, numeracy strategies, and so on). Some of these may well make a difference, but ask for the evidence first.

Another consequence of the path well taken is that innovation can become stifled. Larry Cuban (2005) has evaluated how teaching changes after introduction of NS policies similar to NZs. He identified two polarities in how teachers have taught in schools for the past 200 years – the more teacher-centred methods where teachers control what is taught, when, and under what conditions with the more student-centred where students exercise more responsibility for what is taught and how it is learned. One has much teacher talk and students (pretending) to listen, the other much student talk and teacher listening. Either method is usually defended without substantial evidence, are claimed to apply to all students, and are driven, Cuban claims, by ideology and faith leading to rhetorical struggles over what is best for all students. For example, the New Zealand ideology for teaching mathematics emphasizes the strategies for using numbers, whereas the recent United States National Council of Teachers of Mathematics (2006) report rejected this approach and claimed that mathematics teaching should emphasis math content knowledge, knowing multiplication tables, and how to do division and manage decimals.

Since NCLB, Cuban found teachers spending more classroom time preparing students for high stakes tests and less time on those subjects that are not tested. The test is the goal. There is more teacher talking and explaining, more homework from textbooks, greater use of worksheets, and over training in completing multiple choice items as part of the daily life in schools. Indeed my own 5-year-old in his North Carolina school had to spend 40 minutes a week learning how to fill in bubble sheets in anticipation of the next year’s testing regime. From about 1000 classrooms, Cuban observed, there were more hybrids of student- and teacher-centered methods – in his phrase, there is much “hugging of the middle” – as teachers adapted to the high stakes policies. The concern is that innovation – which occurs outside the middle – can be stifled, and there is a regression to the mean.

Another step on the path is narrowing the curriculum. If literacy and numeracy are the priority, then prioritizing these subjects above others should occur. Dillon (2006) surveyed 15,000 districts in the United States and found that more than 70% had cut back time spent in social studies, science, art,

music, and other subjects to create more time for reading and mathematics – particularly in lower socioeconomic schools. They also cut back on recess time and on the more student-centered activities (e.g., small group work, discussions, learning centres, and portfolios) because such work takes away precious classroom time from standards-based curriculum and test preparation. Watch this path; it is often an early warning sign of the path well travelled and the choice of Near and Far Horizons.

Sooner or later the success of the NS policy is evaluated. A typical finding is that it is very hard to change the mean score of the nation. Elley (2005) reviewed over 160 comparisons of cross-cohort achievement across all year levels in NZ. These tests varied in format from closed to open, practical experiences and writing tasks, administered individually as well as in teams, classes and schools. “Nearly every contrast showed only negligible differences of 1-3%, whether they were based on intervals of one year, four years, seven years or as many as 32 years between surveys, and whether they were assessed in terms of changes in means, or SDs or percentiles. Sixteen of the 31 effect-sizes that could be calculated were .10 or less, only four exceeded .20, all in listening skills” (p. 19). So one conclusion when NS are evaluated is likely to be that the mean achievement does not change and those naughty teachers sabotaged or did not implement NS very well. Or that NS had no impact and should be changed. Before we introduce a policy it should be required to specify the expected outcomes/targets and then evaluate the policy on the merit of these targets, the evidence that the policies lead to achieving these targets, and an agreement to then re-review the policies and the targets. Exhortations that NS will fix the problem demands we know what the problem is – and then we can understand the nature of the evaluation. I predict that NS will barely if at all change the mean achievement scores of the nation.

The politicians, advocates, and officials who promised improvement from their policies then start to become concerned that they do not see the gains promised by their national standards policies. Maybe, they say, the standards were too aspirational, and we cannot have a system that “fails” so many, so it is necessary to take stock, conduct a review, and so on. These can lead to complicity in finding ways to make the policies seem successful. This path include presenting data in ‘interesting’ ways such as using fancy graphs, changing of the standards (in the predictable direction), and higher baying of criticism at the culprits who are causing this lack of success (i.e., schools). I am not saying that we would do this in New Zealand, but I am saying that this is what has happened elsewhere.

Recall the United Kingdom experience. Soon after the SATS (Standard Assessment Tests) were introduced, there were pleas to allow for differences in intake of the students. Why should schools with intakes of more struggling students be compared with those with brighter entrants? Soon after, value added and now contextual value-added using hierarchical linear modeling was used to account for these initial differences. Great work for psychometricians, but there were delays in returning the information, the information was highly complex, and schools had difficulty interpreting these models. This led to fancy internet tools to help interpretation but it was all too late to have much formative effect on teaching and learning.

Then the emphasis in the United Kingdom shifted to progress, which is almost an afterthought. At best, progress gets a fleeting reference but there is no substance on how to evaluate it in practice. The initial plea is “keep it simple”. But evaluating progress is not simple. Students rarely progress in linear fashion, some regress (e.g., start at the curriculum expectation and a year later, be below that expectation), and psychometricians are still grappling with optimal methods to measure progress – what chance do teachers have, then, when the methods are still being developed.

To understand progress, it is necessary to relate this growth to both some content and some normative expectations – progress from one curriculum level to the next only makes sense relative to information about other students’ growth. We know, for example, that growth from Level 1 to Level 2 is far greater

than from Level 3 to 4, and a growth of one level needs to be related to which level the student started at, and whom he or she is being compared with; it also probably needs to take into account the past growth of that student. Imagine saying a student had a 4cm height growth in a year – the interpretation of this change depends on what is expected of other students who started at the same height as this student, their past growth (if they had grown 1cm or 6cm the previous year), other growth indicators (if they had increased in weight by 1kg or 4 kgs), where they started (a tall child's growth of 1cm may be interpreted differently from a short child's growth of 1cm); the measures also need to be based a common metric on a singly underlying continuum (height in centimeters). As Betebenner (2009) has shown, the growth trajectories of students who start at different points of success are greatly different, and thus any individual's growth success or not is not "simple" to calculate or compare. While every student in New Zealand may NOT be expected to achieve the average or even the "aspirational" national standard, it is the case that every student in New Zealand should be expected to demonstrate adequate or appropriate progress from one year of schooling to another. This most critical measure of the success of national standards is not simple to ascertain, far out of the hands of teachers without sophisticated psychometric skills, and likely to be the forgotten part of the implementation of national standards.

As the critics and supporters of NCLB have claimed, a major mistake has been indexing the standards to years of schooling. By such indexing, the implication is that students are taught as if all in the year are of a similar "standard." If we know anything, then the 'years' refer more to the amount of spread in achievement within a year not the stages of growth – that is, year 4 means they are spread by a factor of 4, and year 8 spread by a factor of 8! To then assume that a standard based on a year refers to what the majority of students in that year can achieve is folly and should be resisted.

Other mistakes of policies such as NCLB that we cannot afford in New Zealand include: not acknowledging or effectively addressing the current enormous inequalities in the provision and success of schooling in New Zealand; mistaking measuring schools for fixing them; assuming a theory of action to shame and name will motivate improvement (or change the horizon); undermining rather than expanding safety nets for struggling students; creating debates about student outcomes rather than teacher quality; punishing those who need help as opposed to identifying and resourcing those most in need of help; and basing judgments about schools on test scores (see Meier et al., 2004).

**Reactions to this well trodden pathway**

The path is well known as a result of the experiences of those countries that have proclaimed and implemented national standards. The key message in this paper is NOT that this is the inevitable path that New Zealand will go down; indeed, we will only go down these paths if we do nothing; this is the default path. To prevent walking this well trodden path, we need policy statements that mandate against, warn, and challenge whenever there is evidence that we are going down this path, or better, the development of policies now to ensure we not go down the well travelled path -- before we venture too far into the implementation. Prevention is better than reaction. We need clear, definitive and well developed plans for implementation, agreed warning signs when the wrong path is taken, and agreement as to when to change or abolish the National Standards if they led to perverse outcomes. We also need to celebrate if they are successful at enhancing teaching and learning for all students across the curriculum.

A major step down a less travelled path would be to state the problem that NS is aimed to resolved. It is not overall literacy rates, as we rank among the highest in the world. We do know, however, that too

many students (about 1 in 4) do not gain minimum literacy and numeracy credits in NCEA Level 1; we know that too many students are not progressing as would be desired; and we know there are gaps right across the achievement distribution between Pakeha and Maori/Pasifika students. If National standards are to remedy these problems, we could readily develop "dashboard indicators" which could show, at a glance, student progress in these areas. Similarly, to remain vigilant about student progress in reading and mathematics, we could also pay attention to their progress through the levels in all subjects in the curriculum; pay attention to the progress of those students already at or above the national standards; and pay attention to ensuring that teaching continues to improve. National standards could provide information in all these areas that could be presented in the form of "dashboard indicators" that are aimed enhance teaching and learning for *all* across the curriculum. At a minimum, when the NS were first there should have been publication of the current status of student achievement levels by years and then asked what the target for a national standards policy would be, rather than set an arbitrary wish cut scores (80% plus, and so on) and hoping that the standards could "hit" this target somewhere where desirable.

Parents are another source of information on the extent to which national standards are useful and successful. Assessing parental satisfaction is not sufficient; what is more important is whether parents accurately interpret the national standards information provided by schools. Being satisfied but inaccurate is a disaster, as is being unsatisfied and inaccurate. The aim is not for schools to spend more time creating more student reports to parents, but to spend more time understanding how parents are interpreting this information and using it to become more involved in their children's learning. Even better, if students are to be successful in their learning, in their self-regulation of their learning, and in their own interpretations of assessments, it is optimal that they are involved in this communication to their parents of their achievement and progress relative to the national standards information. The success of national standards could be evaluated in terms of the increase in student-led conferences, for example, and how students approach their learning in ways that are more likely to lead to improved learning.

It is worth noting what parents considered success from a National Standards policy (based on the MoE report on the feedback from the consultation). Parents noted the importance of their children enjoying learning, it should be fun as well as robust, that the "whole child" needs to be developed, as well as their children receiving a good grounding in the basics and beginning a lifelong relationship with learning. Parents asked questions, well worth answering, such as Will the standards be the same everywhere (they may be on paper but are they interpreted the same?), How do we know the standards are set at the right level? Were they set by working teachers or academics (MoE experts) from Wellington? How will the standards be applied consistently particularly with teacher judgments involved in assessments? and Will children be held back in classes if they are below the standard? It is noteworthy that the parental key themes are NOT in support of national standards. Instead, national standards are seen as an implementation mechanism that could achieve the parental goals – and thus these themes could be worthwhile as part of the Far Horizon of an independent review. These themes include: parents wanting specific, honest, and regular feedback on their children's progress in language they can understand, and with early notice of any problems; parents wanting good communication and strong relationships with teachers as part of working together to support their children's learning; parents wanting good information about the curriculum and National Standards, and parents wanting practical assistance and support from schools so they know what to do to support school-based learning in everyday life.

The success of national standards will be related to the quality and dependability of the standards. The current approach of developing standards by committee is not good enough. The glossy, recently published New Zealand literacy and numeracy standards have no data, no evidence, and no evaluation – they are pronouncements without evidence. If there is evidence outside committee contemplations, where is it? Until there is evidence, the standards remain untested and experimental. It seems to surprise New Zealanders that there is a huge literature, multiple methods, books and conferences all dedicated to standard-setting methods. Instead, we "know" the truth, and provided the insiders agree, we seem complacent in our ignorance of the evidence and ignorance of standards-setting methods. Success for standards in New Zealand will be related to using many of these methods of standard setting that are well established outside New Zealand, such as the Bookmark, modified Angoff, and so many others; using these methods would lead to all teachers and students making similar overall judgments and understanding the differences between national standards 1 and 2, 2 and 3, and so on.

The success of national standards could be evaluated in terms of the confidence parents have in the quality of assessment information available to teachers and students. For example, when one local primary school in my neighbourhood ran a parent evening on assessment in the school, more than 300 parents turned up. The parents were given an overall brief of the assessment policy in the school, and then spent 15 to 20 minutes in groups learning about assessments such as School Entry Assessment, Progressive Achievement Tests, Star, running records, e-asTTle, and so on. The major message for the school from this experience was the unbelievable astonishment of the parents that schools knew so much, had such great tools, and had such professional teachers. But they then asked why the school reports were seemingly based on stories, generalized comments, jargon, and lack of evidence. Too many school reports are like this – school reports could be a school's best public relations message, but so often they are their worst, which leads politicians to act on parent concerns about the lack of evidence parents receive on the success of schools. The problem is not that this quality evidence is not available, it is just not shared – and we pay the price of imposed policies when we do not share.

The success of national standards could be related to an increased teacher voice not just in our communities but across the nation. If teacher judgment is critical, then let us hear it and esteem it. Other professions have Queens' Counsels, Royal Societies of Scientists, Boards of Architects, Institutes of Accountants, Colleges of Surgeons, and so on, but teachers have no such professional voice. While teacher unions appropriately are concerned with enhancing the working conditions of teachers (and also do speak on issues of quality teaching and learning), we also need teacher voices to enhance the national debates about the quality of teaching and learning – and be at the table when the policies are discussed and formed. Success can be measured in terms of the emergence, existence, and esteem of our teachers' professional voices – as critics and developers of how the policies are to be implemented and evaluated. This does not mean consultation in the form of attendance at a speech; it means involvement in ensuring that the policies are optimally implemented and evaluated through a well established professional organization and structure.

The success of national standards could be related to breaking down the moats around schools. We currently have 2500 islands with sometimes deep moats around them. We ask schools to be responsible for their own islands and this is a huge task which leaves little energy for building bridges. We ask schools to compete for their student numbers, which means that the bridges they do build need to dismantled for some parts of the year (see Hattie, 2009). We ask 2500 flowers to bloom and wonder why some do not! We ask boards of trustees to be responsible for the quality of the school and when they begin to move down the avenue of failure, we wait until they fail and then ask the same people who got the school to this situation to now get it out – while the learning burns! We have too many

moats and schools working in isolation, and too few collective voices devoted to improvement, enhancement, and sharing of what works best.

If national standards are to succeed, there needs to be an announcement NOW that there will be an independent (a truly independent) evaluation of the effects of national standards on teaching and learning. There may need to be a review of the processes and success of implementation but this is not what is requested (alone). An independent evaluation needs to highlight the claimed goal of the Far Horizon (the effects on the quality of teaching and learning) and not on the Near Horizon (the successful implementation of national standards). There needs to an agreement NOW that the policy will change, or self-destruct, if the quality of teaching and learning for all students across the curriculum is not enhanced. The recent major independent review of the United Kingdom's national standards policy (Alexander, 2009) concluded that the standards agenda is viewed less favourably; not because of opposition to high standards or accountability – far from it – but because of the way the apparatus of targets, testing, performance tables, national strategies and inspection is believed to distort children's primary schooling for questionable returns. The report noted the questionable evidence on which some key educational policies have been based, the narrow focus of policies which treat literacy and numeracy as proxies for the whole of primary education, the disenfranchising of local voice, the rise of unelected and unaccountable groups taking key decisions behind closed doors, the "empty rituals" of consultation, the authoritarian mindset, and the use of myth and derision to underwrite exaggerated accounts of progress and discredit alternative views. In New Zealand, an independent review, announced now, may help prevent such consequences, which often accrue because of the focus on the Near Horizon.

The success of national standards could be measured in the nature of information provided about national standards. This could be in the form of establishing the status of the nation's education each year compared to its status at the time national standards began (in 2009). Success would be judged in terms of more students reaching the different rungs of the standards ladder, and, as important, in the evidence about student achievement progressions (their rate of learning). We can learn from NCLB and demand this information for *all* students (and not just the average student). We will know the national standards policy has failed, however, if it leads to league tables of schools (which is one of the greatest barriers to the success of national standards). We know New Zealand has among the lowest between-school variance in the world, which means that if you take two students of the same ability, it almost does not matter which school they go too. So why create the diversion of a debate about between-school differences, when most of the variance in student achievement is within the individual schools. The success of national standards policy is related to the nature of the debate (success of students and schools), not to the differences between schools. Until the issue of league tables is addressed, it is unlikely that the hearts and minds of teachers will be gained.

The presence of league tables will be the greatest cause of schools teaching to the tests, thus perverting the nature of teaching — and league tables are very hard to remove. Also, they often ignore measurement error, which many researchers have shown is sufficiently large to make the league tables meaningless (Leckie & Goldstein et al., 2009). It does seem ironic that in the United Kingdom, league tables have been promoted as a basis for school choice, but the statistical models used (contextual value added) include school compositional effects which make choice based on league tables meaningless. Leckie and Goldstein concluded that the sophisticated system used in the United Kingdom is highly misleading – it inappropriately uses current school performance as a guide to future school performance, it identifies only a tiny group of schools that can be significantly separated from the national average (or from any other school), and the models do not adjust for prediction uncertainty. As

the United Kingdom experience indicates, the presence of test score league tables could alone be the point of failure for national standards in New Zealand – in terms of failure to implement the standards, failure to develop defensible league tables, and failure to create discussion about teaching and learning rather than between-school differences (and New Zealand has even less between-school variance than the United Kingdom).

There are several alternatives to test score league tables. For example, they may include forming tables based on ERO school reviews (with one-, two-, or five-year periods between reviews), as such tables may provide a more effective overall evaluation of schools than test score league tables (provided ERO gets its evaluations correct). Schools could be encouraged to create their own "dashboards of success" (that is, information on current levels and progress compared to when national standards began), and involve their communities in their development. Such transparent dashboards of success may minimize the effects of the crudeness of league tables. Schools could place more emphasis on independent evaluations and self-reviews of the success of teaching and learning on achievement across the curriculum for all, as these may show the fallacies (or truth) of the information in such league tables.

The success of national standards is also related to the nature and effectiveness of teacher professional development – especially professional development about assessment, overall teacher judgments, quality interpretations of evidence, and moderation of judgments. Failure with national standards is related to schools becoming even more awash with data, the presence of data warehouses (that stores terabytes of data with little if any interpretation and value), and a narrowing of the testing methods in schools. Where is the plan for assessment professional development, what is the budget for professional development, where is the evaluation of this professional development in terms of its effects on teaching and learning. As was noted in the earlier discussion on the implementation of NCEA, the issue of teacher moderation of "overall judgments" will be a necessity – where is the plan to get in front of this issue and not wait for moderation to be the next front page scandal? It will be so much easier to place a test in front of a parent when saying "your child is well below standard" than providing a "teacher overall judgment." This abrogation of responsibility by schools, or this naïve pressure from a parent will be a major disaster for the nature of teaching and learning (and probably explains why so many teachers in other countries have asked for more testing!). We would not want only the test results from a doctor; we would want his or her interpretation, recommendations, and actions. Further, resolving the moderation issue is neither simple nor trivial, but we know so much about how to reduce it, we know effective moderation models (and ineffective models), and we can provide resources to schools to minimize this problem.

The success of national standards will depend on teachers accepting information that comes to them from previous teachers (and across the school divides – from contributing to intermediate to secondary schools). The agreement of teacher judgments across these divides will need to reflect a common conception of progression in national standards, a reduction in wasted time at the start of each year as teachers re-assess where the students are at before progressing upwards in the curriculum, and building a stronger professional community of talking and trusting each other's judgments.

The success of national standards will depend on parents understanding school reports and not merely on satisfaction with or independent reviews of the quality of school reports. Clarity and satisfaction are not the aims; the aim is dependable *interpretation*. This will require the community of schools to similarly interpret the levels and sub-levels of our curriculum (and to hold a common understanding of "above", "at", "below" and "'well below"); otherwise parents will be confused. To achieve this will require much professional development for teachers to ensure they have the requisite moderation skills

and the capability to interpret assessment results, as well as much teacher debate and the provision of resources.

The success of national standards will depend on ERO not only dependably and earlier identifying the schools that need moving in a new positive direction in their effects on student learning, but also being consistent around the country in asking about the evidence a school has about how its students are performing and progressing and asking what the school is doing about that evidence. Any ERO evaluation that emphasizes data over interpretation should be seen as inadequate, and any parent who does not become convinced of the value of teacher judgment over the test score should have concerns about the quality of the teacher and school. Any evaluation across the country as a whole could be in terms of the increase in schools receiving ERO reviews only every three or more years (as this indicates fewer schools needing frequent reviews because of poorer quality), and a light sampling of schools summary data could suffice to show the system the effects of its policies.

The success of national standards will depend on the standards NOT becoming the de facto curriculum, and on the appropriate use of assessment. We may need to hold the feet of the promoters of national standards to the flames relating to their many pronouncements that there will be no national test(s). It will be a sign of failure is this occurs (although if it occurs, it will surely be claimed, as has been the experience in other countries, that the failure is "because" the standards were not effectively implemented!).

The success of national standards can be measured in terms of the nature of the debates in the community and across the nation. Success will be measured by discussions about how impressive the assessments are in our schools, how rich the interpretations that are provided to all, how capable our students are in monitoring their own learning, how self-review of each school's contribution to teaching and learning (not league tables), and how successful New Zealand teachers and schools are compared to the rest of the world. We know from the annual Gallup polls that about 80% think the quality of schooling in the country is going backwards; and 80% of the parents think the quality of schooling their children receive is excellent. Discussion on the latter would be a better measure of success than discussion on the former

Most critically, success will depend on the choice of horizon. Is the success of the national standards policy to be evaluated in terms of how it enhances teaching and learning; or is the success of the policy to be evaluated in terms of how teaching and learning is changed to implement the national standards? We need to answer this NOW – we need an agreement that there will be an independent evaluation (i.e., an evaluation NOT overseen by the Minister or Ministry) of national standards in terms of how they have had an effect on the quality of teaching and learning not just in literacy and numeracy but across the curriculum for *all* students. We also need agreement that there will be a "self-destruct" button that will get rid of national standards if the policy is not shown to make the difference to teaching and learning across the curriculum.

**Conclusions: The whirlpool effect**

Any change to government policies, priorities and regulations always creates opportunities. Given that the language in these national standards pronouncements is lofty, aspirational, and missing much detail, what a wonderful opportunity there is to exploit the whirlpool of innovation. The current literacy and numeracy policies offer many opportunities for schools to implement strategies that we know work and can truly make the difference. We could take the opportunity to instill capability in their (schools, teachers, students) own assessment of their learning, to continue with assessment for learning, to

develop teacher capability in the assessment of students, to involve communities in schools, and many more. We have a levels-based curriculum and almost 20 years of experience in interpreting and using it; we have a recent curriculum which we are required to implement, and thus national standards should be subservient to this mandate. This whirlpool of opportunities should be the major focus of debates in and across schools as the implications of the national standards policies unfold.

The standards road is well travelled. It seemed apt that a major review of the United Kingdom system was released the week the national standards were launched in New Zealand (Alexander, 2009). The messages were clear – implemented poorly, national standards can have perverse effects; implemented well, they can make a difference. The report warned particularly about making "spectacular assertions", such as claiming that the state of primary education was at a "low state", that testing of students drives up standards, or that SATS (standards-based tests) are the only way to hold schools to account. The report warned about starting formal schooling too soon, asks for high standards in all and not just some curriculum areas, recommends that assessment for accountability be uncoupled from assessment for learning, and says that the debate should move from *whether* schools and teachers should be accountable (they should) and should instead concentrate on *how*. New Zealand must not go down the path hat lead to these conclusions. We know the danger signs now, thus the implementation must be mindful of them if and when they occur. At minimum, we need a guarantee that there will be a similar independent review of the goals and implementations of New Zealand national standards in about two years' time. We cannot afford to expend our energies and goodwill solving the wrong problems, fixing the perverse directions, and not knowing whether policies, when implemented, are making a positive difference to the quality of teaching and learning *across* the curriculum for *all* students.

Success in implementing national standards in literacy and numeracy should be evident by seeing:

- evidence of current levels and progress and targets as to desired levels and progress
- the use of defensible standard-setting methodologies and not standards set by committee
- that the national standards do not become the de facto curriculum
- that the "games" (increased retention, accommodation, blaming students) are not mistaken for evidence
- that innovation is not stifled and the curriculum not narrowed and standards dumbed down
- all students advantaged by the policies and no-one is excluded from teaching and learning
- students appropriately challenged and no teaching as if all students in the one year are similarly challenged
- clear practical methods for evaluating progress
- esteem given to moderated, overall teacher judgments (rather than to test scores)
- the meaning of "levels" and "progress" similarly understood by all teachers and students consistently throughout New Zealand
- teachers trust the assessment information that travels with each student from previous teachers and schools
- parents understand the information provided to them (by schools, and, even better, by the students) and parents esteeming the quality of assessment and interpretations by teachers (and students)
- the equity problems in New Zealand remedied, and failing schools identified and remedied earlier
- improvements and evidence-based ideas shared freely and quickly around New Zealand schools
- "dashboards" for school progress and success developed to offset a narrow focus on only one part of the curriculum

- the absence of debates related to “between-school” differences (e.g., league tables) and increased sharing across schools
- ERO asking about the quality and consequences of school judgments (and never asking only for the data)
- no development of data warehouses and anything but light sampling to evaluate effectiveness across schools
- the emergence of highly accomplished teacher voices on a national scale
- a mandate for a truly independent evaluation of national standards in terms of its effect on teaching and learning and the desired targets of levels and progress within two years, and agreement to “self-destruct” the policy if it is not working.
- increased professional development in assessment (overall teacher judgment, levels, progress, quality interpretation of evidence, moderation of judgments, developing student assessment capability)
- an increase in the success of schools and in the confidence of parents and students in their schools
- evidence (not just claims) that the Far Horizon is that national standards are aimed at enhancing teaching and learning across the curriculum for all students

Michael Barber (2007) was the architect of the United Kingdom’s SATS and testing program to enhance literacy and numeracy (and not, note, the quality of teaching and learning). He was very successful as he attended diligently to the process of implementation. He noted the four keys to implementation: understanding the degree of challenge, the quality of planning for implementation and performance management; the capacity to drive progress in implementation; and the timetable for delivery. We have glimpses in NZ of each right now, and thus there is an opportunity for schools to contribute and even control these implementation processes. He advised “gentle pressure, relentlessly applied” and regarded success as primarily related to clear and realizable targets; sharp accountability; good real-time data; best practice transfer; transparency; management against implementation; capacity to intervene when necessary; and incentives to reward success. We can apply these implementation processes both to the Minister and the Ministry in New Zealand, as well as in individual schools. Success towards the Far Horizon may accrue as long as the goals are agreed and in harmony, and the implementation is on the path we wish to travel.

We may have one of the best systems in the world, but we cannot defend a system with one in ten schools deemed failing; one in three students failing the minimum Level 1 numeracy and literacy; every school competing to devise the optimal systems, sometimes too alone. Nor can we defend a system that decides on one policy (national standards in literacy and numeracy) and not get it right – both right in the policy, the standards, the implementation, and its evaluation in terms of the correct horizon. It may be necessary to have edicts about the chosen horizon; otherwise schools may have the best intentions of aiming for the Far Horizon, only to feel the pressure to reverse the relationship based on feedback and pressures (e.g., league tables) from parents and others, particularly the media to aim for the Near Horizon. A major success will be dependably identifying struggling schools and students earlier and seeing the national standards policy truly making a difference to them. National standards offers the most wonderful opportunities for refreshing and reinvigorating an already top of the world system, but it could be the most disastrous policy formulated if it turns our attention to narrowing, testing, league tables and diverting attention to between-school rather than within-school differences.

We have much to do.

**References**


Alexander, R. (2009, June). The primary curriculum: An alternative way forward. Westminster Education Forum Keynote Seminar: *The future of the primary curriculum.*

Barber, M. (2007). *Instruction to deliver: Fighting to transform Britain's public services.* Methuen, London.

Betebenner, D.W. (2009, April). *Growth, standards and accountability*. Paper presented at the American Educational Research Conference, San Diego, CA.

Cuban, L. (2009). *Hugging the middle: How teachers teach in an era of Testing and accountability*. NY: Teachers College Press

Dillon, S. (2006, March 26). Schools cut back subjects to push reading and math. *The New York Times*, p. A1

Elley, W. (2005). On the remarkable stability of student achievement standards over time. *New Zealand Journal of Education Studies: Te Hautaka Matai Matauranga,* 40, 3-24.

Fullan, M. (2003). *Change forces with a vengeance*. London: Routledge Falmer.

Hattie, J.A.C. (2009). Tomorrow's schools – Yesterday's news: The quest for a new metaphor. In J. Langley (Ed.). *Tomorrow's Schools: 20 years on*. (pp. 121-134). Auckland, NZ: Cognition Institute.

Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Association, 172,* Part A, 835-851.

Meier, D., Kohn, A., Darling-Hammond, L., Sizer, T.R., & Wood, G. (2004). *Many children left behind.* Boston, MA: Beacon Press.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence*. Reston, VA: National Council of Teachers of Mathematics.



Citation. Hattie, J.A.C. (2009, October). *Horizons and whirlpools: The well travelled pathway of national standards*. Working Paper from Visible Learning Lab, University of Auckland. Located on Cognition Research Trust Website, www.xxxx.xxxx