

# Occasional Papers

2003

## Assessment in the Real World – What Teachers Actually Do in the Classroom

*By Mary Hill*

*In the 1990s the reintroduction of national standards in education through a national curriculum required schools to be accountable for measuring and reporting on their learning outcomes. Primary teachers became caught up in assessment frenzy.*

*How did teachers adapt to constantly changing practices and what impact did it have for learning?*

### **Investigating primary teachers' practices in the 1990s**

The aim of my research was to discover why primary teachers in New Zealand were conforming to particular managerial practices, such as recording large amounts of assessment information on checklists when they had reported these as being unhelpful. In contrast to these checklisting practices, almost all the teachers stated that they believed assessment practices should be used mainly to inform their teaching and improve learning and achievement.

For this study I drew on data gathered from interviews, observations and documentary analyses. I worked with teachers in two Waikato primary schools to investigate their assessment policies and practices. As well I interviewed 20 primary teachers throughout New Zealand to see if their policies and practices resonated with those in the two case study schools. I interviewed 12 teachers in depth over two years and observed three of these teachers each for a full week in action in their classrooms.

## Primary Teachers' Assessment Knowledge and Practices

In explanation, although half of the twelve teachers stated in the first interview that they decided what assessment was appropriate in their classroom, most of the teachers indicated that assessment expectations were driven from above.

We have school guidelines that state (what) we have to assess in all areas and that, in turn, comes from Ministry requirement... so really it comes from the top, comes from the Ministry and then it filters down through the board, the principal, the senior teachers, down to the ordinary teacher.

In the end the teachers just go into automatic mode and go along and tick, tick, tick and at the end of the day they don't really mean anything anyway ... and I find that a lot of the things that are being assessed at the moment are the things that are easily assessed and therefore they are sort of almost driving what's being taught because that's what ERO seems to be wanting.

Teachers found it hard to articulate their ideas about learning theory and assessment. They dismissed as jargon the classification of assessment into formative and summative, and frequently talked about how they recorded and reported information when asked about their assessment practices. In addition, they reported an increase in the amount of recording of assessment information that was required. They were concerned about their ability to implement school-wide systems that would be manageable, lead to better learning and enable them to monitor progress against the national achievement objectives: in short, reconcile the different formative and summative purposes of assessment as highlighted earlier. The results from a wider national sample of 85 schools confirmed that most schools were wrestling with the same challenge.

## Classroom Assessment

There was some significant differences in the ways teachers went about their assessment practice. I therefore decided to observe, in depth, the classroom practice of three teachers who had described contrasting assessment, recording and reporting practices, each for a week. Analysis of their classroom practices and their responses to further questions confirmed that teachers seemed to have three differing approaches to classroom assessment.

### A) Assessing Separately from Teaching

Most teachers appeared to focus their attention on students' ability to achieve the objectives set for a specific unit of work. Typically, these teachers planned from the New Zealand curriculum documents, choosing appropriate achievement objectives and sometimes reducing these to several specific learning outcomes from one or two levels of the curriculum. They mainly saw teaching and assessing as separate activities, describing instances of assessment as reasonably formal, planned events. One such teacher told me that:



While we're all doing fairytales they'll be all meeting different needs within the fairy tales unit.... So I suppose I'll assess on that, can this group do blends, is this group able to use a contents page, is this group able to tell me about the characters in the story.

Most of these teachers employed a three or four point scale to record each student's achievement of each objective. One teacher explained:

I use a triangle system. (One side means) yes, we've done something about it, and here (two sides means) they're on the right track and (all three sides means) yes, they've definitely got it.

This generally occurred after teaching and often towards the end of a week or a unit of

work. The priority for these teachers was to gather evidence about each student's progress against the achievement objectives in the New Zealand curriculum. These teachers - commonly listed the achievement objectives directly from the curriculum as headings in their data/roll books. Although these teachers said that they usually did not use this assessment information for formative or diagnostic purposes, except at times to regroup students, when asked to explain how they made teaching decisions they gave details such as:

I think in our heads we have a sort of map and a timeline. I can sort of see how they progress and where they might be starting from and where I'd like to see them go.

Hence, these teachers framed assessment for formative purposes as 'teaching' not assessment. Interviews and observation with one teacher who worked in this way revealed that while she found checking each student against the achievement objectives 'overbearing', she did this because she was expected to. And she explained that when another teacher (usually more senior than her) appraised her, this was an aspect that would be checked. She explained that each level within the school hierarchy would check that the next was performing the functions allotted to it to a certain standard. She explained that the expectations are set out in the school's policy.

We have a minimum (of running records) that we have to do for the year. We have to take story and handwriting samples every term and it's expected that we have a roll book in the junior school, where we collect data on an ongoing basis on units done and on reading, maths, story writing ... It's expected that we cover all areas frequently.

Thus the systems set up to monitor the progress and achievement of students disciplined the teachers as well.

## **B) Headnoting**

Other teachers tended to rely mainly on their memories of what students could do. Some also jotted notes during teaching. Although referring to the achievement objectives in their planning, these teachers had meagre systematic records. In fact, they stated that keeping records often interfered with teaching. These teachers said they tended to focus more on 'teaching' than 'assessing' or recording. Several of these teachers described their assessment as part of teaching, focussing on individual learners and moving them along in different aspects of their learning as necessary:

I work with the children on the floor going over the work, it might be decompositions, subtractions and I discuss and work with them until they can do it, then they go away and do practice. I'll get their book afterwards and find out how they have gone. If they understand it then I can move them on, if they don't I'll get them back and see them. So that's a way of finding out where they're at and that's on a daily basis and that's extremely effective because you cater for the student's needs.

These teachers generally kept samples of each student's work as a record and for reporting

purposes. But about once a term or twice a year, they set assessment tasks and checked off progress against the achievement objectives. Again, this was recorded by most with a checklist system.

### C) Integrating Formative Assessment

A small number of teachers demonstrated how they used both planned and unplanned interactive assessment during teaching and learning activities and recorded information about each student's progress towards meeting specific objectives as well. A teacher in the junior school explained her use of such a document to me in this way:

So it's with me, nice and close to me whenever I want it... I know the document (her data book) so well, ... that as the children are doing their work, I can quickly just note down in it what's happening or what they need. I have it on hand. I couldn't work without it. I need to refer back every so often.

These teachers had designed methods by which they tracked the progress of individual students through a large number of specific concepts and skills, and acted on these results, mostly during teaching. One teacher explained her approach to learning and teaching thus:

It's a need, I need to know before I can move to the next stage, have they got that? So it's got to be done, ongoing the whole time. Otherwise how would I know what to teach next?

## A Short History of Assessment

Prior to the establishment of a national compulsory system of primary education in 1877, assessment practices in colonial schools were knowledge-centred and, at times, quaint. The language of 'standards' entered the vernacular after 1877, when national standards were set for primary students who, after their initial years of schooling, were actually put into classes called Standard 1, Standard 2, and so on until Standard 6 (now Year 8) when students sat a national 'Proficiency' examination. Thus, in Standard 1 for example, a student would be expected to achieve the following 'Reading' standard: *Sentences composed of words of one syllable, and common words of two syllables, to be read intelligently.*

Unsurprisingly, teachers began teaching to the test, and it was not uncommon for students to learn, by rote, prescribed books. In the 1920s and 30s, however, progressive teachers focussed on the 'natural' development of students and challenged the ideology of national, uniform, prescribed standards. This challenge co-existed, rather uneasily, with standardised testing that purported to measure students in terms of externally referenced norms.

While standardised testing did not disappear it became gradually less important. Between the 1950s and the 1980s the emphasis shifted towards more descriptive, 'child-centred' and diagnostic assessment, with a focus on individual student's abilities and how they might be

fostered. The pendulum swung towards attempts to systematically record student progress through such devices as running records and writing conferences.

It wasn't until after the restructuring of education in the 1990s that concepts of accountability, based around marketisation and managerialism, re-established themselves, and 'national standards', started struggling anew for dominance. Much of this was driven by *Tomorrow's Schools 1989* which made schools self-managing and individually responsible for reporting their students' achievement, and by the introduction of a new levels-based national curriculum in 1993. The reintroduction of national standards in education through a national curriculum brought in outcomes, levels and requirements for schools to each be accountable for measuring and reporting on their learning outcomes.

### Contemporary research into assessment

During the 1990s international research into assessment increased dramatically. Much of this research focussed on primary education in the UK, the USA, Canada, South Africa, Australia and New Zealand. The findings indicated that the field of assessment research was rapidly expanding, complex, and, at times, confusing. Taken as a whole, however, the studies in the early 1990s revealed that assessment activity in primary education was focused very much on performance (outcomes) rather than on learning (process and conceptual development).

Two major reviews of the research on assessment (Crooks, 1988; Black and Wiliam, 1998) indicated that this outcomes emphasis was likely to undermine students' motivation and achievement. In fact, Black and Wiliam stated that:

*There is a firm body of evidence that formative assessment is an essential feature of classroom work and that development of it can raise standards. We know of no other way of raising standards for which such a strong prima facie case can be made on the basis of evidence of such large learning gains.*

Formative assessment provides feedback to teachers and students about on-going progress in learning, usually during learning and teaching action. It directly influences the quality of students' learning experiences and, thus, achievement. In contrast, summative assessment communicates the nature and level of students' achievements at various points in their schooling. Summative assessment has at least three purposes:

- assessment for reporting achievement and progress;
- assessment for certification, used to summarise achievement for selection and qualification;
- assessment that judges the effectiveness of teachers, educational institutions and of the system as a whole.

Formative and summative assessments are not two distinct types of assessment with accompanying distinctive practice. They are overlapping, conflicting and often confused. But it is clear that teachers needed to understand the critical importance of regular, relevant and specific feedback to learners, and 'most vital of all' (Crooks, 1988, p. 470) that assessment must emphasise what matters over what is manageable.

The priority for these teachers was to use the assessment information they gathered to plan for curriculum differentiation, for groups as well as individual students.

I observed how one such teacher checked to see if significant behaviour she was noticing was actually 'new' for that particular student. I saw her call a group together towards the end of a session in which she had been roving and recording as she taught. Sending the rest of the class to work with the second teacher in the room, this teacher focussed specifically on teaching the group to use the beginning sounds of words when they were not sure how to proceed in reading. This only lasted for a few minutes, but demonstrates how she used her records to calibrate her teaching. The teacher described here was not only working on improving students' learning across the curriculum as she taught, but she was also using assessment for formative, diagnostic and summative purposes, simultaneously. The checklists she designed came from her knowledge of how students learned particular curricula rather from the curriculum achievement objectives. In other words, this teacher had been able to set out the map and time line she had in her head and use this as a guide for teaching and assessment.

### **The Impact of Accountability Demands on Teachers' Practice**

From even the small amount of data presented above it is apparent that New Zealand teachers have felt the accountability demands placed on them in self-managing schools. All teachers in this study described how they gathered evidence to meet the summative requirements of syndicate and school policy and the National Education Guidelines (NEGs) and the National Administrative Guidelines (NAGs). Some clearly managed



**Teachers deserve help to re-establish the central place of formative assessment in learning and teaching.**

## In Search of Quality, Manageability and Validity

*Talk of template planning and reporting along with a centralised system of analysing student achievement data against prescribed targets has, understandably, been greeted with scepticism by school practitioners. The 'selling point' for these changes (as outlined in the Education Amendment Act 2001) is that it will be simpler and more manageable for schools, but the variables thrown up by crunching huge amounts of uncontextualised data could produce all sorts of misinformation, entirely unconnected with ongoing student learning and school programmes. School level sampling for achievement trends and patterns, on which professional judgements are made, could well be a better alternative. Colin Tarr reviews some literature on this approach.*

### Some current issues facing schools

In her longitudinal survey on 'Tomorrow's Schools' Wylie (1999) found 90% of teachers surveyed said that the amount of assessment they did had increased over the previous three years. Timperley et al (cited in Wylie, p 142) suggests, while teachers are doing a great deal of assessment, they may be doing more than they need, and not making adequate use of what

to balance teaching and accountability demands while others had more difficulty. Many focused almost entirely on checking progress against the achievement objectives. The widespread use of checklisting systems suggests that teachers found them manageable and that they helped meet the policy requirements. In the climate of accountability which required teachers to 'monitor students against the achievement objectives' (Ministry of Education, 1993b, p.11) and checked up on them through performance appraisal, school review and through the Education Review Office visits, this is, maybe, unsurprising.

The main drawback with most checklisting systems, however, is that they are not geared to guide students in how to improve their learning nor can they diagnose individual strengths and learning needs. This checklisting approach may not, therefore, facilitate the very learning that is desired. The Education Review Office itself, in a study of 118 schools in late 1998 identified that many schools were using this sort of system, which ERO called 'quasi mastery assessment'. ERO criticised these checklisting systems as of little use. It seems ironic that these systems were developed by schools largely to meet the NEGs and the NAGs and often in response to perceptions of ERO reports that required them to monitor progress against the national achievement objectives.

they collect to evaluate and improve teaching programmes because their focus remains on the individual child.

Much of the increased assessment work schools undertake is to provide information for school-wide accountability purposes. The National Administration Guidelines – NAGs (Ministry of Education 1999) require schools to identify and report on students who are not achieving, at risk of not achieving, who have special needs or aspects of the curriculum that require particular attention.

In a search for quality, manageability and validity, I wonder whether a school level sampling approach could be, validly and reliably, used to assist in meeting these NAGs requirements and the worthy principles espoused in *Tomorrow's Standards*.

### Overview of some commentary and research

Sutton (1998) said one of the considerations affecting a school's self-review strategy is manageability. The attraction of concentrating on quantitative data (eg test results), she says, is the relative ease with which it can be gathered and analysed. Qualitative data is harder, slower and more expensive to collect hence there is a systemic aversion to scrutinising students' work in depth, observing them at work in the classroom and undertaking longitudinal monitoring of learning development. Sutton maintains the challenge is to make such data gathering manageable without undermining its quality therefore *sampling for purpose* should be the aim.

Many of the teachers in this study tended to keep most of their ongoing assessment information in their heads. These teachers explained that they used learning progressions in their mind when teaching, but that they recorded progress and achievement at the end of units or less frequently – generally when required to do so to meet school policy demands.

Research suggests that it is possible to use evidence collected for formative purposes for summarising achievement providing that the different uses are understood and that the teacher knows how to use the information in their teaching as well as how to report it for accountability purposes. However, the prominence of checklisting following the reforms suggests that many teachers have found it very difficult to prioritise assessment for formative purposes. These checking approaches seem more consistent with performance and outcomes than with educative purposes. Even though teachers are using criterion-referenced assessment, the criteria are often used as a hierarchy of performances to be checked off, graphed and reported. The real teaching and learning intentions and criteria may exist only in the minds of these teachers.

The school-wide policies and quasi-grading systems I observed had been developed

Well-planned sampling, she argues, could generate high-quality information and provide a pointer for further enquiry.

Irving (1997) agrees that *'selected samples of sound classroom assessment information can provide the basis for aggregation and analysis in order to show overall school effect'* (p.14). He is cautious, however, about the types of information that constitute evidence. Irving maintains that a good deal of classroom assessment information is directly, and desirably, formative and diagnostic (eg observation) and may not be suitable for aggregation. He suggests this rich and valuable information will not be suitable for pure accountability purposes. He maintains, however, that classroom-based information can have a high validity and could be useful for deriving summaries that profile school progress and achievement over time.

Gipps (1994) is cautious about the use of sampling at a school level. She confirms it is an appropriate methodology to reduce the stakes for national and district level accountability purposes and may be acceptable in large schools to get valid school level performance and accountability data. She maintains it will not work in smaller schools where there are not enough students to provide an acceptable sample size.

In discussing national monitoring Flockton (1999) asks educationalists to consider the implications of assessing every student at a particular year level compared to a random selection of students. Where assessment of all students is required, he argues, it is likely to be in the form of paper and pencil tests, multiple choice and machine marked. New Zealand's *'National*

in response to the National Education Guidelines and were reinforced through feedback from ERO visits. Syndicate and school-wide policies definitely influenced teachers' checklisting practices. For example, quasi-grading systems that encouraged teachers to dot, slash, cross (or construct triangles) against achievement objectives and/or learning outcomes at each level of the curriculum were reported as policy in most of the schools visited.

## **Maximising the Potential for Assessment to Enhance Learning**

In the last ten or so years it has been teachers, first and foremost, who have had to deal with competing assessment theories. Many teachers expressed frustration at recording assessment in line with school policy while, at the same time, operating their teaching according to the learning progressions they have in their minds. Teachers deserve help to make connections between their own tacit craft knowledge and the curriculum objectives, to re-establish the central place of formative assessment in learning and teaching, and to find non-intrusive ways of recording and reporting outcomes.

*Education Monitoring Project* (NEMP) uses a carefully designed 3% sample of students at Years 4 and 8 and uses a wider range of assessment methods to obtain a richer coverage of skills and processes. Flockton advises that to obtain a national picture of student achievement, it simply is not necessary or advisable to test every student.

The Ministry of Education's Research Unit team (1998) confirm gathering good (i.e. valid and reliable) data is an essential element in any school's endeavours to evaluate its own performance. They say that including all members of the target population in an investigation will, often, not add significantly to the trends in the information obtained. Sampling is suggested as a suitable and viable option in that there is less information to analyse. *'When sampling there are no clear cut rules for deciding on sample size. Generally speaking, however, the larger the target population the larger the sample that is needed'* (p. 35). A well-selected sample that is representative of the target population allows for generalisation to occur.

In order to mitigate the cost of assessment Shepard (cited in Sanders and Horn (1995)) suggests sampling of students and grades in key subjects through the use of exemplary assessments. This 'trade off', however, 'places severe limitations on the uses to which the assessment results can be put.'

Indeed Sanders and Horn say that if such a suggestion were implemented, the resulting data would be practically useless for assessment of educational entities, in that while it reduces costs it does so at the expense of information on individual performance. They maintain that

Presently, the Ministry of Education's professional development programmes in assessment and tertiary level teacher education courses seek to address these goals. But syndicate and school-wide policies also heavily influence teachers' practices. A more fruitful approach to learning could be school assessment policies that encourage teachers to provide the feedback crucial to student learning. Teachers, and in particular, principals as the leaders of schools, need to understand how assessment can inform teaching and learning. They need to know how to use assessment tools and other rich assessment tasks formatively, as well as for measuring achievement for reporting. They need to be informed about how to implement school policies that use assessment wisely.

Rather than setting targets and reporting achievement against the achievement objectives in every curriculum area, schools could assist teachers to focus on priorities, such as aspects of literacy and numeracy. They could do this by developing policies that require information about only very important areas to be collected and reported, as well as used for improving teaching, so that teachers can put their emphasis on using formative interactions in their teaching rather than on reporting on achievement targets.

School guidelines about what assessment data teachers need to collect should be

information must be generalizable if the progress of a cohort of students, on the basis of limited data, is to be logically accomplished through time-honoured statistical means.

Having a conceptual understanding of good assessment principles is vital for educators seeking to undertake school wide assessment tasks. McMillan (2000) says professional judgment is the foundation for assessment and that evaluation is concerned with judging the merit and worth of data applied to a specific use or context.

### **A discussion on these findings**

A key motivation for the use of sampling is manageability. Sutton (1998) and Irving (1997) maintain that well planned sampling is possible at a school level without compromising the quality of information obtained. The Ministry of Education's research team (1998) agrees that when a sample is well selected it allows for some generalisations to be made about trends and patterns occurring within the target population.

Sanders and Horn (1995) say that generalisations about the progress of student cohorts must be able to be made from the information gleaned from the sample. If they can't be the sample design is flawed and of very limited value.

Assessment information is no more than a snapshot in time on the learning journey. Irving (1997) cautions about using classroom-based formative assessment information for school accountability purposes although he does maintain that such information can have high validity

clear and not dominate the classroom programme. Periodic monitoring of important indicators of learning should be enough to demonstrate progress. Schools should encourage the use of assessment tasks that provide information that can be used to diagnose strengths and needs, and provide achievement data. For example the following can provide information that can be used for both purposes:

- the 6 year old diagnostic survey,
- items from the Assessment Resource Banks (ARBs),
- the National Education Monitoring Project (NEMP),
- the forthcoming national exemplars,
- Assessment Tools for Teaching and Learning (asTTle) assessments.

Often, in order to meet managerial demands, school leaders are instrumental in developing policies that increase the amount of testing, checklisting and recording. Even though some school leaders read research reports and realise that formative assessment is vital for enhancing learning, their lack of knowledge and understanding about what formative assessment actually is can lead to management practices that hamper its implementation.

and with careful use could be used to derive some valuable information about student progress over time.

Researchers have mixed views about the value of sampling as a method for national assessment systems. Gipps (1994) addresses sampling in the context of national assessment. She suggests that sampling at a school level is inappropriate given that sample sizes, in most cases, will be insufficient to provide valid data from which generalisations could be made. Flockton (1999) argues that sampling for national trends allows for a wider, richer array of tasks to be undertaken that generate highly useful information. The debate appears to centre on 'either/or' scenarios. Either large-scale quantitative data gathering (i.e. standardised test marks) or a sampling approach that, its proponents argue, provides opportunities for the gathering of richer qualitative information. It seems the crucial element in the sampling approach is in the assurance of validity and reliability of process and product.

Researchers agreed that sampling at school level could be a useful tool, if designed carefully enough. Thus it seems it is an approach worthy of some further scrutiny and investigative trial.

## Conclusion

Mugo (1999) describes various types of sampling approaches in her paper '*Sampling in Research*'. She explains a judgement sample as one obtained according to the discretion of someone familiar

Other culprits driving the summative over the formative use of assessment information are software systems that are used for analysing and reporting assessment information. Most require checklisted information or test results to be entered in order to produce graphs, charts and tables. The checklisting approaches revealed in this research can be exacerbated by purchasing such systems and by requiring teachers to use a quasi-mastery system of achievement objective checking to make them work. While computerised systems appear to save time by producing graphs and charts for reporting on achievement targets or well-presented reports for parents, teachers are often expected to enter the assessment data electronically themselves, often on slow hardware systems. This takes more time away from planning and teaching. Such systems should be purchased and used judiciously.

One of the strongest influences on teachers' assessment practice, apart from the way in which teachers are required to collect and record assessment information, is the way teachers themselves are assessed. Appraisal and performance management practices can reinforce assessment practices that enhance learning as opposed to those that engender testing and checklisting. Again it is important that school managers, including principals

with the relevant characteristics of the population. Purposeful sampling is described as an approach where information rich cases are selected for in-depth study over time. These approaches can be stratified to illustrate characteristics of particular subgroups of interest and facilitates comparisons between different groups.

If a manageable and reliable school-based sampling programme is found to be of use such an approach would not replace ongoing individualised assessment programmes. It would, however, be a valuable additional tool to help the school see what trends are apparent and what areas might require further in-depth probing and review in order to improve their children's learning. Information, perhaps, a 'datacube' and data crunchers in Wellington wouldn't be able to provide?

## REFERENCES

- Flockton, L. (1999). *School Wide Assessment: National Education Monitoring Project*. Wellington. New Zealand Council for Educational Research.
- Gipps, C. (1994). *Beyond Testing: Towards a theory of educational assessment*. London. Falmer Press.
- Irving, J. (1997). *School Wide Assessment: The Big Picture*. Wellington. New Zealand Council for Educational Research.
- McMillan, J. (2000). *Fundamental assessment principles for teachers and school administrators*. Practical Assessment Research and Evaluation. Retrieved April 21, 2001 from the World Wide Web: <http://ericae.net/pare.getvn.asp?v=7&n=8>
- Ministry of Education (1990). *Tomorrow's Standards*. Wellington. Crown Copyright.

and curriculum coordinators, understand the influence of formative assessment on enhancing learning and how to design teacher appraisal policies and practices that promote formative over summative uses of assessment.

## Conclusion

National policies, and school expectations, policies and procedures are extremely powerful influences in determining teachers' assessment practices. Teachers and school leaders need to be confident about assessment so that they can make informed decisions about their classroom practice, so that they can prioritise formative strategies over accountability ones and so that they can see how summative and standardised assessment can work for, rather than against, learning. School leaders could assist teachers in raising standards through assessment by constructing school assessment policies that do not pressure teachers into gathering summative data at the expense of using formative assessment practices.

While teachers would appreciate help to improve their existing assessment practices,

- Ministry of Education. (1999). *Sharpening the Focus No.1*. A pamphlet series on the revised National Administration Guidelines. Wellington. Crown Copyright.
- Mugo, F. (1999). *Sampling in Research*. Trochim's Centre for Social Research Methods. Cornell University, New York. Retrieved March 28, 2001 from the World Wide Web: <http://trochim.human.cornell.edu/tutorial/mugo/tutorial.htm>
- Sanders, L. & Horn, S. (1995). *Educational Assessment Reassessed: The usefulness of standardised and alternative measures of student achievement as indicators for the assessment of educational outcomes*. Education Policy Analysis Archives, Vol. 3, No. 6. Retrieved April 21, 2001 from the World Wide Web: <http://epaa.asu.edu/v3n6.html>.
- Sutton, R. (1998). *School Wide Assessment: Improving Teaching and Learning*. Wellington. New Zealand Council for Educational Research.
- The Ministry of Education Research Unit Team. (1998). *Some Helpful Steps for Schools to Carry Out their Own Research*. The Research Bulletin, No. 9, October 1998. Wellington, New Zealand. Ministry of Education.
- Willis, D. (1992). *Educational Assessment and Accountability: A New Zealand case study*. Education Policy Journal, Vol. 7, No. 2 p 205-221. Wellington. Taylor and Francis Ltd.
- Wylie, C. (1999). *Ten Years On: How Schools View Educational Reform*. Wellington. New Zealand Council for Education Research.



**Colin Tarr** is the national vice president of NZEI Te Riu Roa and principal of Pukeatua School. In 2000 he was director of the NZEI, NZPF, NZSTA team facilitating information workshops on the revised National Administration Guidelines for the Ministry of Education. His recent MEd study is to do with the evaluation of educational programmes.

it seems clear that increasing accountability measures further will do little to improve standards. Recent measures requiring schools to set annual student achievement targets and report to the Ministry of Education on the extent to which these have been met (Ministry of Education, 2002) may well jeopardise progress towards balancing formative and summative assessment. Although attempts have been made to keep the stakes low by not requiring information in a way that can be used for league tables of schools, there are still inherent dangers in the national surveillance of school results (Hill, 2002). The very act of requiring the reporting of targets and variance between these and actual results, may be perceived by schools and teachers as raising the assessment stakes in the same way as teachers in the research changed their practices due to their perceptions of ERO's expectations. Even teachers with the best knowledge and understanding possible cannot resist the performance imperative when the stakes are raised (or even perceived to be raised) for purposes of comparison. If we are serious about raising standards in education the evidence points strongly towards the development of formative practices (assessment for learning) in classrooms.

The continuing challenge is for New Zealand teachers to construct policies and procedures that promote teachers' use of assessment to raise standards and improve

student achievement over the need to continually provide accountability data. To achieve this:

- teachers need to be well educated about learning, teaching and assessment;
- policies, both at the national and school levels, must prioritise learning and conceptual development over outcomes;
- and national monitoring of standards in education should not raise the stakes for comparison between schools or teachers but, in contrast, increase the emphasis on investigating trends and patterns of achievement through initiatives such as the National Education Monitoring Project.

## REFERENCES

- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-74.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58 (4), 438-481.
- Education Review Office. (1999). *Student assessment: practices in primary schools* (Education Evaluation Reports 2). Wellington, New Zealand: Education Review Office
- Hill, M. (2002). The Education Standards Act and Improving Student Achievement: Flipsides of the same coin or chalk and cheese? *NZ Principal*, 18(2), 5-9.
- Ministry of Education. (1993). *The New Zealand Curriculum Framework*. Wellington, New Zealand: Learning Media, Ministry of Education.
- Ministry of Education (2002). *Sharpening the Focus*, 7, 1-4.

Note: This is an edited version of Mary's Hill's paper. The complete text is available on the NZEI Te Riu Roa website [www.nzei.org.nz](http://www.nzei.org.nz)



**Mary Hill** is a senior lecturer and Co-ordinator of the BTchg (Primary) degree at the University of Waikato. She investigated primary school teachers' assessment practices in self-managing schools for her PhD. Mary has an intense interest in learning, teaching and assessment and teaches courses at undergraduate and postgraduate levels in these areas. She is presently part of a team of researchers at Waikato undertaking the National Sampling Study for the Ministry of Education's Curriculum Stocktake. She is also involved in an international research project investigating changing views about professionalism in primary schools. She can be contacted at the School of Education, University of Waikato.