

Connecting Theory and Practice in National Monitoring Assessment

Lester Flockton

“The National Education Monitoring Project (NEMP) is part of the Government’s strategy for obtaining information about achievement standards and the quality of education.”

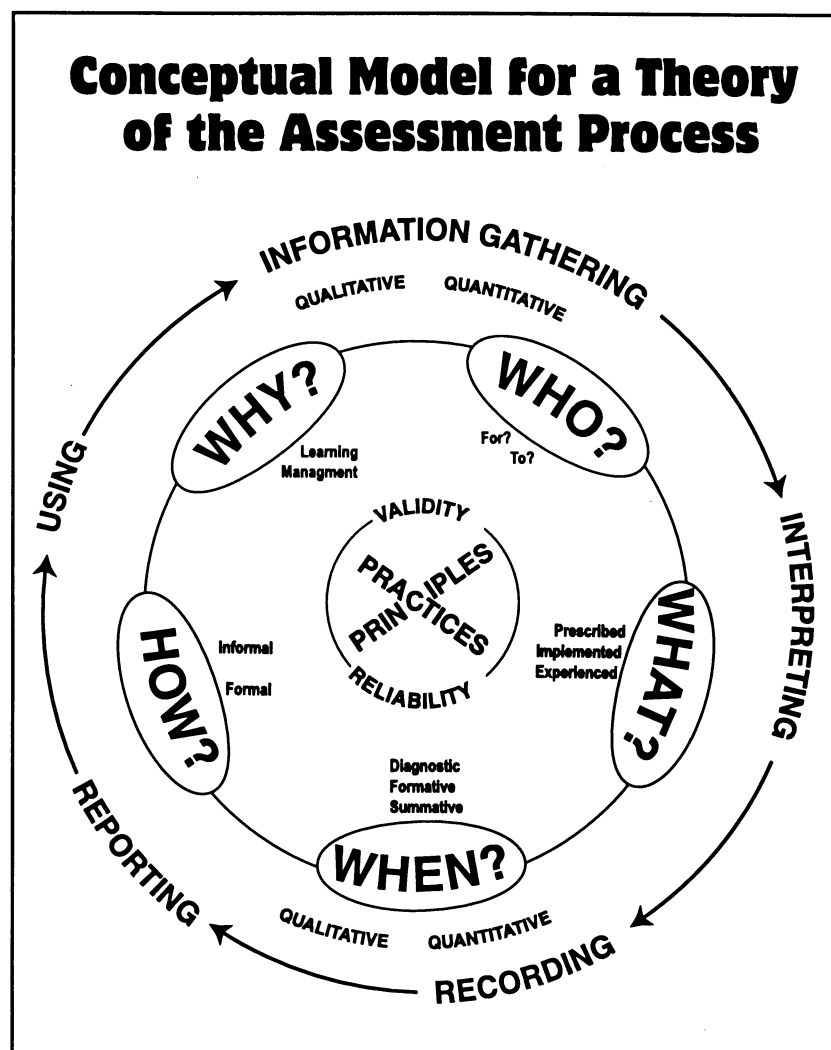
The National Education Monitoring Project (NEMP) is part of the Government’s strategy for obtaining information about achievement standards and the quality of education. Underpinning this official purpose is a rationale that focuses largely on the consequences of assessment for teaching and learning, while not overlooking the importance of the technical quality of the information.

The rationale is sensitive to the connections between assessment theory and the many practical

demands on teachers for implementing the national curriculum, which includes monitoring and assessing student achievement.

If connections between theory and practice are to be meaningful, useful and more than hope, then it is vital that the form and meaning given to a theory of assessment links directly into practice.

The word assessment itself is used and understood in many ways and contexts, and can have quite different meanings for different constituencies.



Lester Flockton
Co-Director
Educational Assessment
Research Unit
University of Otago



In the current climate, assessment and accountability are seen by many as driving forces for helping all students achieve high standards of performance.

In NEMP, assessment is understood and treated as a process rather than an event. It is a process which involves three critically interrelated actions: (i) the identification of *important* learning outcomes; (ii) the collection of both qualitative and quantitative information related to those outcomes from a wide range of performance tasks, observations, and other techniques, then (iii) making information on the outcomes available in ways that help to determine, illustrate and guide the improvement of individual, group or programme performance.

Implicit in this meaning of assessment are important notions of purpose (who and what the assessment is for), methodology (and the effects of methods used), and an understanding that good assessment practice derives substantially from good teaching practice.

If it is accepted that a theory of assessment is about a contemplative *scheme* of something to be done, or an exposition of general principles, assumptions or methods, then such a scheme, in outline, will need to represent critical dimensions of the design and operation of an assessment programme such as national monitoring.

Four Pillars of Learning

Education should be built on four pillars:

- **Learning to know** is the basic learning that provides a foundation for life-long learning together with the possibility of in-depth work on selected subjects.
- **Learning to do** emphasises the capacity to apply knowledge in a variety of situations.
- **Learning to live together** addresses a deep need; and understanding of others and their history, traditions and spiritual values – and this contributes to achieving a spirit that recognises our society’s essential interdependence and the need to work together *harmoniously and constructively*.
- **Learning to be** emphasises the spiritual side of humanity and sees personal fulfilment as an ultimate aim. It also implies the need to develop qualities of imagination in children and adults as life long learners.

Learning, the Treasure Within, UNESCO, 1996

Among the various aims we consider important in education, two are especially so. We would like our children to be well informed – that is, to understand ideas that are important, useful, beautiful, and powerful. And we also want them to have the appetite and ability to think analytically and critically, to be able to specu-

late and imagine, to see connections among ideas, and to be able to use what they know to enhance their own lives and to contribute to their culture.

Elliot Eisner

Wynne Harlen, director of the Scottish Council of Educational Research, implores educators to recognise that learning needs to be about helping the learner “gather meanings directed towards bigger ideas about our place in and experience of the world about us”. This is a good summation of the beliefs held by those who shape the practice of New Zealand’s national monitoring.

What, then, is the basis for your own professional beliefs and practice? Try yourself on this multiple choice test item:

What is the basis of what guides your professional thought and practice:

- (a) What I know and believe from study and experience.
- (b) What we know and believe from collaborative endeavour.
- (c) What others say we shall know and believe.
- (d) What a technocratic reductionist insists upon as good practice.

KEY PRINCIPLES FOR ASSESSMENT

Use and produce trustworthy information.

The words “reliability” and “validity” typically spring to mind! Unless assessment results have sufficient technical quality, they are prone to giving shaky or misleading information about the work of students, teachers and schools. However, “life is not multiple choice”. As learners for life, students need to be able to apply what they know to approach problems, create solutions and communicate ideas. The issue here is that what is assessed needs to be “worth teaching for”! That is at the core of validity, but often in conflict with attaining clean-edged reliability.

CENTRAL IDEA

The core or central idea in the scheme outlined here emphasises that what is done in assessment should derive from a few well considered principles. These will be principles that support a coherent and reasoned belief system about the meaning and purpose of education itself; a belief system which is capable of being both internalised (understood) and externalised (practised) by those who are most closely connected to the student in the educative process.

BELIEFS

Speaking to a conference audience in New Zealand last December, Elliot Eisner referred to classical Greek philosophy which held that "beliefs must be true – otherwise they are opinion", a rationale which is sullied by the

"For educators, the importance of constantly checking practice against beliefs is an essential professional action."

reality of an empirical world which is imperfect – just as the processes and products of assessment are also inevitably imperfect. Regardless, the importance of having and articulating beliefs about the purpose of education and the place of assessment is a fundamental professional responsibility for all educators at every level. But beliefs need to be more than opinion or supposition – they need to be *reasoned* standpoints that can be explained and defended with some clarity and passion as well adding meaning to the content and practice of teaching and learning. If beliefs are widely **shared** they will provide a common and fruitful foundation for the enactment of educational policy and practice at every level, whether it be the classroom, the school or the system as a whole. In reality, demonstrable harmony between beliefs and practice is more attainable at the individual school

level than at a system or societal level, or indeed between individual schools and the school system as a whole. This is unfortunate, because disharmony or dislocations in belief systems can lead to actions and reactions that typically result in counterproductive conflicts and power politics.

For educators, the importance of constantly checking practice against beliefs is an essential professional action. Like curriculum and schooling in general, assessment policy and practice should be intrinsically connected with our beliefs and visions. If we don't regularly review policies and practices for their connections with what we believe, then we jeopardise the "moral purpose" of our vocation.

The following thematic and contrasting examples of beliefs about the aims or purposes of education can help stimulate reflective or contemplative thought about how various views fit not only with one's own thinking and reasoning, but also with views on practices that are seen, followed or supported.

- **People as "rational utility maximisers"**

I see the school as being nothing more than a finely focused factory and, like any other factory, we import the product. In our case it's students, and we then, like any other factory, add value to that product and we eventually aim to produce a product that is both in demand and appropriately skilled to meet the expectations of the market place.

Alistair Rivers writing of *New Zealand Schools*, 1996.
Quoted in Snook, "Reflecting on the Rules of Enterprise,"
Education Review, July 1996.

- **New Zealand Curriculum Framework, 1993**

The New Zealand curriculum will enable students to "develop (their) potential, to continue learning throughout life, and to participate effectively and productively in New Zealand's democratic society and in a competitive world economy".

"Students are asked to talk about their ideas, to construct, experiment, observe and explain as well as sometimes producing written answers."

As is increasingly being recognised, the preoccupation with reliability has necessarily tended to lead to a concentration upon what is more easily measurable and a relative, if not absolute neglect of higher level intellectual skills such as thinking and those more affective qualities and meta-cognitive skills which are now recognised as being much more difficult, if not impossible, to measure using traditional multiple choice tests or even written examinations.

Editorial, *Assessment in Education*, vol. 1, No. 1, 1994

National monitoring faces the challenge of developing and following assessment methods that fit with a belief system about what is important for teaching and learning. We therefore attempt to emphasise approaches that focus on a broad range of important learning outcomes, build on good models of teaching practice, and provide useful



feedback to support improvement and motivate performance. These intentions give a rationale for what is broadly termed “performance assessment”, an approach which seeks to set tasks within meaningful and authentic contexts.

Because performance assessments usually require that students’ work is marked by teachers exercising their professional judgements, they do pose challenges for achieving acceptable technical quality of information. NEMP uses procedures built around marking schedules (rubrics). Cross-marking and teacher collaboration so that basic agreement on scores falls within tolerable limits. To achieve substantial consensus on standards and expectations, processes of discussion, guidance and exemplification are essential. But the central consideration, and advantage, is that teachers are at the core of critical decision making.

Use the best available assessment practices

National monitoring considers two interrelated dimensions of assessment practice: the methods or approaches used, and the effects on students of the methods and approaches.

Many traditional testing methods have serious limitations for assessing higher level skills like thinking, reasoning, creating and communicating, so national monitoring prefers to use approaches that allow students to produce responses that have “something of meaning”. Students are asked to talk about their ideas, to construct, experiment, observe and explain as well as sometimes producing written answers.

Good performance assessment taps complex thinking and/or problem-solving, addresses important disciplinary content, invokes authentic or real-world applications and uses tasks that are instructionally meaningful. Furthermore, because they require students to construct a unique answer, performance assessments typically are

scored by humans, exercising judgment, rather than by machines.

Joan Herman, 1999

Consider also the effects that assessment practices can have on students, and consequently their results. If tasks are set within contexts that are remote from their life and learning experiences, or at difficulty levels which are unrealistic, students are likely to become discouraged and unmotivated. If rigid time limits are imposed, performances will more likely reflect what can be done *within* time, rather than what can be done *with* time. Timing might be appropriate on the race track, but its usefulness as a constraint on demonstrating what is known or what can be done is highly questionable.

Very few national monitoring tasks impose rigid time limits.

Focus on change over time

Consideration of the reasons for having information about gains or progress helps to clarify the particular purpose for which assessment is conducted. At a national or system level, the purpose is largely concerned with getting a dependable picture of what students know and can do at particular points in time and from one period of time to another. This provides a basis for evaluating the impact and suitability of curriculum in relation to important learning outcomes, and the effectiveness of teaching and resources. At a school or classroom level it will be to do with the effectiveness of programme design and delivery. For the individual learner, it will help identify current needs and future learning goals.

Any measure of progress needs to recognise that it takes time to make genuine progress in the development, understanding and retention of knowledge, skills and ideas. Moreover, progress is not an all or nothing dimension of performance. It might be fast, moderate or slow depending on the *rate* of learning that has occurred over time.

Assess a broad range of outcomes

The national monitoring project upholds the importance of a balanced curriculum which spans all learning and major content areas. This is consistent with a belief system which values learning for living in the present and future. Assessment tasks need to communicate what is important to learn, and if those who are tested care about the results, they’ll be motivated to perform.

Involve practising teachers

Teachers through their professional training, development and sustained experiences in the real world of the classroom have valuable and valid insights and understandings about the nature of curriculum and its relevance to the diverse groups of students they teach and communities they serve.

Development towards good practice is most likely to arise when professionals themselves perceive what is on offer as being credible, sensible and worthwhile. If they are

“Since national monitoring is fundamentally committed to contributing in the best possible ways towards the improvement of teaching and learning, it values a strong partnership with teachers.”

partners in the guided development of policy and practice, they are more likely to understand, support and accept its influence and direction. Having a strong sense of commitment and ownership is fundamental to achieving genuinely meaningful professional growth and improvement of teaching and learning.

Since national monitoring is fundamentally committed to contributing in the best possible ways towards the improvement of teaching and learning, it values a

strong partnership with teachers. Increasingly, studies are showing the instructional value of well conceived and designed assessment programmes and tasks. Not only can they communicate what is important to learn, but they can also provide good models for the kinds of teaching activities and processes teachers should use in their classrooms.

Use assessment information for improvement

Reported information can act as a catalyst for improvement, but improvement also requires identification and acceptance of need, effective strategy, and the resolve and resources to carry it through. By having substantial involvement of practising teachers in all facets of the work of national monitoring, the real and perceived qualities of validity are increased and acceptance of responsibility for professional action is heightened. The project supports the belief that extensive teacher participation strengthens ownership of successes and challenges, and that ownership is at the core of the professional action necessary for advancing teaching and learning.

We need to be cautious, however. Good assessment practices or programmes may create the will to change but not the capacity to do so unless they are combined with effective professional development programmes.

Research shows that teachers and principals take new performance assessments and the goals they represent seriously and often try to incorporate new pedagogical practices into their teaching. Teachers attempt to engage their students in the kinds of activities they see embodied in the assessment. However, in the absence of sustained professional development, these classroom innovations will likely lead to superficial changes in practice that have little impact on student learning.

Joan Herman, 1999



**INTERLATIONSHIPS: WHY?
WHO? WHAT? WHEN?
HOW?**

Why?

Before elaborating on principles and methods, any theory of assessment needs to address the question, why is the assessment being done?

Of the three major purposes for assessment (learning, management and certification) NEMP serves both learning and management.

Learning – to inform and guide curricular and teaching decisions.

Management – to provide information on the relative performance of subgroups (girls/boys; Maori/non-Maori; large schools/small schools; urban/rural, etc.).

Arguably, all assessment should have the aim of supporting and improving learning, and indeed most assessment information does have that potential to a greater or lesser extent. However, it is the way the information is *used* that is critical to assisting the improvement of learning.

If, for example, worrying gaps are revealed in the performance of one subgroup relative to another, chastisement from the foot-light parade contributes little towards closing the gap. Instead, the gap itself becomes the issue.

At a professional level the more constructive response lies in identifying the underlying causes for the gap (and more often than not they are many and complex), searching for ways to substantially close or eliminate the gap, trying them out, then regularly monitoring the effectiveness of the intervention. Just as the doctor's diagnosis on its own doesn't cure an ailment, neither does assessment fix learning deficiencies. It can only be a form of temperature taking. Its value lies in telling us where the student is at in relation to expectations of achievement.

Who?

Identifying who should be involved in an assessment

“Just as the doctor's diagnosis on its own doesn't cure an ailment, neither does assessment fix learning deficiencies.”

programme requires consideration all constituencies that have an interest or a part to play. When asking who national monitoring is for, the simple answer is everyone. Policy makers need dependable data and feedback, just as the general community is entitled to have access to information about the performance of students across the education system as a whole. Ultimately, however, the real benefits and greatest connections should be with learners, whether those learners are whole schools, groups of teachers or individual students. NEMP embraces these learners at every point in the design, organisation and implementation of the project. Each year panels of educators come together to advise on task development, teachers assist with producing task ideas, groups of senior tertiary students do some of the marking, dozens of practising teachers work for extended periods on administering and marking tasks, and hundreds of schools and their students participate in the programme. Then every school is sent copies of reports which describe the tasks and the results.

Programmes for system-level monitoring of educational outcomes typically monitor at up to three levels of the compulsory education system. The most junior level is generally about age 9 because by that age most students are able to read and write reasonably well, follow instructions, and have sufficient confidence to respond to the rather special requirements of the monitoring programme. The most senior level is placed at or near the end of compulsory education. The middle level is usually about half way between these two extremes, at about 12 to 13 years of age. New Zealand only monitors the first two of these levels (Year 4

and Year 8) so we don't have the advantage of what would clearly prove a highly interesting tracking of trends right through the school system in a consistent and meaningful way.

Most national, state or provincial monitoring programmes base their monitoring on assessment of relatively small samples of students. They do so because it is cost effective. A very dependable picture of student performance levels can be obtained from a well constructed national sample using a rich range of assessment tasks. Sampling has advantages of avoiding some of the behaviours associated with “high stakes” assessment programmes where, for example, teachers might be inclined to teach to the test. Because whole cohort testing requires that tasks be administered by hundreds of different teachers in their own schools, the accuracy of results is threatened, and yet further demands are placed on classroom teachers for often quite limited benefits. Furthermore, mass testing typically relies upon routine paper and pencil tasks which can be administered to a whole class simultaneously then machine marked. This drastically reduces the range of outcomes which can be assessed and gives rise to a serious mismatch between the learning that is believed to important and the sort of learning that is tested and reported.

National monitoring in New Zealand has very successfully adopted a system of light sampling which each year involves about 125 randomly chosen schools and 1500 randomly chosen students at each of the year 4 and year 8 levels.

What?

We need a system of evaluation that permits us to monitor the educational growth of all children toward the ultimate objectives of education, not simply toward those of a limited number of school subjects.

Coffman, 1993.

The road to hell is sticking to the syllabus.

Eisner, Dunedin, 1999

National monitoring aims to assess the broadest possible range of achievements using the New Zealand Curriculum Framework as a blueprint for coverage of curriculum areas, skills and attitudes. The framework recognises the broad purposes of education and sets parameters for a balanced curriculum. However, to construct a notion of student achievement according to the objective-upon-objective detail prescribed in official curriculum statements would be to deny the need for a dynamic perspective of what students could know and be able to do in a changing world. It could also be simplistic in recognising how and where learning actually takes place. Value is added from multiple sources!

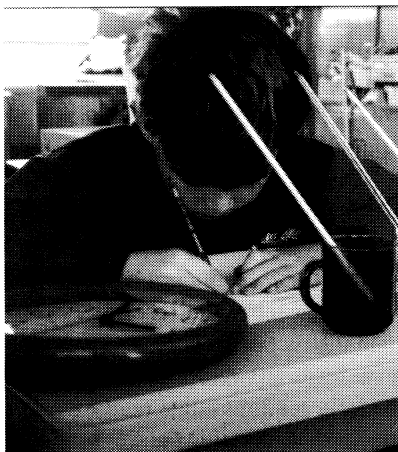
“It was found that a substantial number of students who did poorly in classroom mathematics assessments were actually quite capable mathematicians who did poorly in class because they could not adequately read instructions or present findings in writing.”

In reality, curriculum operates at three levels: the *official or intended* curriculum is mandated by the state on behalf of society at large; the *implemented or planned* curriculum is what teachers actually understand and take from official curriculum and other sources; the *experienced or attained* curriculum is what students really do experience and learn whether from classroom instruction or from endless opportunities or interactions in the playground, community or home. What is clear is that boundaries are not neatly drawn around *what* students might learn, or *where* they can learn it.



Since a legitimate role of national monitoring is to inform the ongoing interpretation, redefinition and resourcing of curriculum, it avoids specific pegging of individual assessment tasks to individual achievement objectives. National monitoring needs to take a long-term perspective, because curriculum areas are to be monitored at four yearly intervals for many years to come. If the monitoring tasks are too heavily based on the detail of curriculum statements, many of the tasks may become inappropriate when those statements are revised. So national monitoring walks the tightrope between failing to take adequate account of the current curriculum and placing too much emphasis on it.

Furthermore, there are serious pitfalls when performances on a few objectives are held to represent achievement across the wider domain of learning - when clearly such generalisation would be very misleading! National monitoring therefore operates from the level of



the major content areas or “strands”, then within each of these it focuses on the big pictures rather than numerous little pictures. For example, in the writing strand in the English curriculum the project follows the advice of expert curriculum panels in identifying the important skills or abilities that represent achievement within the domain. These “big pictures” are then formalised in NEMP assessment frameworks and provide the basis for developing and selecting tasks.

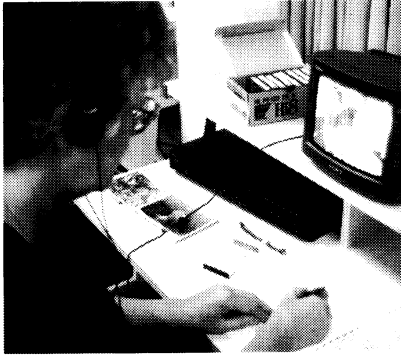
When?

In order to cover a broad range of curriculum outcomes, yet do so in sufficient depth, national monitoring is spread across four years, and then the same cycle is repeated in subsequent four year periods. There is no clear international pattern relating to the time intervals between assessments in each cluster of outcomes (eg. Mathematics outcomes). In the United States, some areas are assessed every two years, others at intervals of four years, and others on a somewhat irregular schedule. In general, marked changes in national performance levels are unlikely to be observed over periods of less than four or five years. However, as the interval between assessments increases beyond five years, the magnitude of curriculum change (whether in theory or practice) can become so large that the use of common assessment tasks across two consecutive periods becomes controversial and potentially misleading.

In a sense, the timing of assessment is also related to its formative and summative uses. For national monitoring these uses can be summarised thus:

Formative or Feedback Uses: ie. “What are the next steps in learning – how do we take them?”

- Rich experience for students taking part in the project.
- Teacher administrators and markers thinking about their own goals and practices.
- Reporting impact– task ideas and



performance findings used by schools and individual teachers.

Summative Use

- A panoramic snapshot in time of what students have achieved to date.
- A summing up in time in relation to what becomes reported (“public”) criteria which derive from curricular/assessment frameworks.

How?

National monitoring uses a combination of assessment approaches including one-to-one interviews with experienced teachers, group co-operative activities, hands-on tasks arranged in a series of stations, and more conventional tests and questionnaires where several students attempt the same tasks simultaneously. Each approach endeavours to set conditions which allow and encourage students to respond to the best of their abilities.

In our early investigations we became convinced that national monitoring could not adequately be conducted through the use of questionnaires or tests administered simultaneously to a group of children. We were strongly influenced by what we saw during our visit to the Toronto *Benchmarks* project. In this project children were assessed through individual interviews conducted by experienced teachers. The children undertook hands-on tasks, and in many cases their efforts were videotaped for later analysis. The one-to-one format allowed students who would not have been able adequately to understand written instructions to listen to the

instruction, and have them clarified if necessary. The interviewing teachers were also able to get students to amplify their responses or to probe the students’ understanding through follow-up questions. It was found that a substantial number of students who did poorly in classroom mathematics assessments were actually quite capable mathematicians who did poorly in class because they could not adequately read instructions or present findings in writing. These observations, together with important findings of the *Learning in Science Project* at the University of Waikato, gave strong persuasion for one-to-one interviewing as a data gathering approach in national monitoring.

From the outset it was also fully appreciated that much learning takes place in groups, and that interpersonal attitudes, communication skills, and group problem solving processes should appropriately be assessed through the use of group tasks. There are few cases around the world where group tasks are regularly used in system monitoring programmes, so considerable exploratory work is demanded of New Zealand’s project, accepting that the approach has very high validity.

A major trend in system monitoring programmes has been the greater use of performance tasks, often involving hands-on use of materials and apparatus. While science led the way with performance assessment, New Zealand’s national monitoring extends this authentic approach to assessment into every curriculum area. Indeed, it is difficult to imagine how many important learning outcomes could be investigated without the use of performance assessment.

A format which is quite widely used for system assessment involves students working around a series of “stations” in which tasks have been set up. At each station the student is given instruction, tackles the task, and records answers. This approach generally requires some reading and writing skill, but since national

monitoring is designed to allow teachers to be readily available to provide appropriate assistance to individual students, this is not a major drawback.

CONCLUSION

National monitoring is substantially about what children know and can do, and the teachers who administer the tasks, mark and use them. The project provides evidence against public criteria, but we avoid making judgements on students’ achievements. That is the role of the widest possible community of interest. The conceptual model for a theory of the assessment process followed by NEMP has been translated into a design which aspires to best practice at every point. However, learning is not uncomplicated, teaching is not uncomplicated, the curriculum is not uncomplicated –and nor is the practice of assessment uncomplicated.

Assessment is an inexact matter and can never be an exact one. There are many reasons for this, the most obvious being that we cannot ever know what is inside the head of a person and we must judge learning from what the person can do in particular circumstances. This means that we do not know what they can do in other circumstances; the generalisability of assessments is limited.

Wynne Harlen, 1994



References

- Boughton, D., Eisner, E.W., Ligtvoet, J. (Eds) (1996) *Evaluating and Assessing the Visual Arts in Education*; New York, Teacher College Press.
- Harlen, W. (Ed) (1994) *Enhancing Quality in Assessment*; BERA Policy Task Group on Assessment, London, Paul Chapman Publishing Ltd.
- Herman, Joan L. (1999) *The State of Performance Assessments*, AASA School Administrator Online.

