



The primary purpose of assessment is to improve students' learning and teachers' teaching as both student and teacher respond to the information that it provides.

NZC p.39

Introduction

1. Curriculum Direction
2. Managing Implementation
3. Curriculum Fluency
4. Dimension 1: Vision
5. Dimension 2: Principles
6. Dimension 3: Values
7. Dimension 4: Key Competencies
8. Dimension 5: Learning Areas
9. Teaching/Pedagogy
- 10. Assessment**
11. Putting it all together

Assessment and reporting have been among the most challenging matters for schools to grapple with over the period of the curriculum documents 1995-2009. There are plenty of experts who would tell us what we should do, and judges to make pronouncements over what we do, but many don't have to do it themselves. Each school needs to develop its *own* expertise. This requires good understandings of the nature of assessment, its possibilities, and its limitations.

The revised NZC is a time for every school to take stock. A time to rationalise assessment practices rather than magnify them still further. If our practices don't serve the best interests of teaching and learning, then they don't deserve too much of our time.

The intention of this set of slides is to give schools a strength of information for examining, reviewing and developing their assessment position and practice.

Collection and Use of Assessment Information in Schools

Education Review Office, March 2007

Key interpretations from the messages in this report:

- The report says many schools are doing assessment that has limited value.
- Teachers need opportunities to develop their judgments about the usefulness of the various tests they are provided with.
- The best assessments are skilled, expert professional judgments from a range of information. Over-reliance on any one assessment or test is potentially harmful - and the results potentially misleading.
- There are some over-stated claims that certain tests improve teaching and learning. These sorts of claims create mythologies.

Localising Assessment

Ten (as it happens) key questions to guide the school's practice

1. Why assessment? What are our purposes and priorities?
2. What are the expected benefits of our assessments and assessment systems?
3. Can or do our assessments and assessment systems cause any harm?
4. What is expected of us? Where does it say that? What does it say? What scope is there for interpretation and localisation?
5. What are the criteria for quality assessment?
6. How will we know that our quality criteria are being met and maintained? What evidence will we have?
7. What are we going to assess? What are the connections with our curriculum?
8. How much assessment is sufficient for our purposes?
9. Do our students have a part to play in assessment? What is it?
10. What recording system will we use, and how will learning be reported?

The New Zealand Curriculum



Assessment *Principles for Practice*

The revised New Zealand Curriculum provides the opportunity for schools to examine and review their assessment and reporting practices.

The question: “What shall we assess and report?” is best answered by the priority goals the school sets for its students. If clear goals are not established, then the basis for assessment and reporting is absent. But it is not just a matter of *what* to assess. It is also a matter of *how* assessment should be understood and conducted.

The NZC provides solid advice to inform the shape of a school's assessment and reporting practices.

Assessment for the purpose of improving student learning is best understood as **an ongoing process** that arises out of the interaction between teaching and learning. **It involves focused and timely gathering, analysis, interpretation and use of information that can provide evidence of student progress.** Much of this evidence is “of the moment”. Analysis and interpretation often taken place in the mind of the teacher, who then uses the insights gained to shape their actions as they continue to work with their students.

NZC p. 39



The NZC also provides what amount to a set of principles to underpin assessment practice.

Effective assessment

- Benefits students
- Involves students
- Supports teaching and learning goals
- Is planned and communicated
- Is suited to the purpose
- Is valid and fair.



NZC p.40

Teachers obtain and interpret information from a range of sources and then decide to use the evidence it provides, based on professional judgment. Teachers can have most confidence in assessment when the evidence comes from more than one assessment.

Many schools already have statements that set out principles to guide their assessment practice. They are also used as the reference points for regular review of the effectiveness and quality of school practice.

Maori Hill School has such a statement in its curriculum plan. The school's "objectives" for practice were decided together by the teachers, and shared with parents and the Board of Trustees.

Maori Hill School

ASSESSMENT OF STUDENT PROGRESS

PURPOSES *Refer also, NAG 1 (ii)*

- To determine students' capabilities and learning needs.
- To monitor the effectiveness of programmes for individuals and groups of students.
- To have up to date information about student progress and achievement for teaching and reporting.

OBJECTIVES

1. *First priority is given to assessing student achievement in literacy and numeracy and then to breadth and depth of learning related to the needs and interests of students, the nature of the school's curriculum defined in this plan, and the scope of the National Curriculum Statements. see. NAG 1 (ii)*
2. *Assessment information is used to identify students and groups of students who are not achieving, who are at risk of not achieving, who have special needs, and aspects of the curriculum which require particular attention. Ref. NAG 1 (iii)*
3. Assessment methods benefit student learning and teaching.
4. Assessment is a continuous and integral part of the teaching-learning process.
5. Assessments are made against planned objectives for student learning in each curriculum area, with a focus on the major learning objectives in this curriculum plan.
6. Assessments reveal individual student successes and weaknesses in relation to (a) age/stage expectations and (b) their own achievement capabilities
7. Methods of assessment are manageable, efficient and use a range of informal to formal data gathering procedures. No one test on its own is a sufficient measure.
8. Qualities of strength of reliability (consistency) and validity (usefulness and meaningfulness) are sought in all assessment data.
9. Teacher professional judgment based on consideration of evidence from a range of information is the basis for determining each student's achievements and progress.
10. Assessment information is regularly used by teachers to give students specific feedback and guidance.
11. Assessment practices include discussions among teachers, parents and students about progress, attainment, effort and priorities.
12. Written records are kept up to date, give a balanced account of each student's performance in each learning area, and are maintained according to management requirements including consistency with the school's Privacy of Information Policy.

School Self-Review of Assessment Policy and Practice

Key Questions

1. Are our purposes clearly stated, and are our practices clearly related to those purposes?
2. Do we have well considered, unambiguous principles (policy cornerstones) to guide practice?
3. Is practice publicly reviewed against declared purposes and principles?
4. Does our curriculum documentation provide the direct reference for what is assessed - and how it is assessed?
5. Do we focus on “big picture” assessment rather than lists and lists of “small picture” objectives?
6. What strength of reliability (consistency) and validity (usefulness) does our assessment information have?
7. Do our assessment practices feed into ongoing teaching and learning, and identify student achievement and progress?
8. Are our systems efficient and not unduly time consuming?
9. Does our assessment recognise the underlying impermanence (fragility) of much assessment data?
10. Do we recognise that all assessments are **approximations** (not absolutes) drawn from **samples** of what students can do in response to the nature of the task.



Assessment

Its nature and purpose

Assessment is a *process* of:

1. selecting/collecting relevant evidence *in relation to ... (the teacher's/school's priority goals for student learning)*.
2. interpreting/analysing relevant evidence *in relation to ... (the teacher's/school's priority goals for student learning)*.
3. using relevant evidence *for purposes of ... (supporting students' learning ... reporting students' achievement)*.

**Is the formative-summative
separation a worthwhile or
necessary one?**

**“All assessment has a formative
component”.**

Jeff Smith

The “prepositional divide”

~~Assessment *for* learning
Assessment *of* learning~~

In recent years we have delighted in separating out functions of assessment as being either formative or summative, “for” or “of” – the prepositional divide. It is time to move on. Schools may be better served by seeing the function of assessment as being ***for learning and reporting***. Arguably, the two functions should be interdependent, one informing the other, even though they each have their distinctive approaches (e.g. formative is teaching/learning interactive; summative is “summing up” valid information on student learning).

Assessment *for* learning *and* reporting

Assessment *for* Learning *and* Reporting *“formative in a summative climate”*

1. Being explicit about what is being learned, and what successful learning looks like
2. Questioning and dialogue - allowing time for students to consider and discuss, often collaboratively, their answers to oral questions
3. Giving feedback that helps students know where they are at and how best to move forward
4. Informing teaching programmes and practice
5. Using self and peer assessment, where students are increasingly able to make good judgments of their own work through understanding and using criteria for quality performance.
1. Referencing to “big picture” learning goals identified in the school’s curriculum
2. Having qualities of dependability (reliability and validity)
3. Addressing “base” learning (literacy, numeracy) and “rich” learning (learning competencies, attitudes)
4. Communicating clearly and unambiguously
5. Telling where the student is at in relation to expectations appropriate to that student (progressions of learning)
6. Making judgments from multiple sources of information (not one test)
7. Describing particular successes, *and* priorities for moving forward
8. Having the student engaged in the process.

Assessment 'Literacy'

conversing theory and policy in practice

The strength of our assessment practice depends on the strength of our assessment 'literacy'. This includes knowing, understanding and being able to discuss:

- What assessment means
- Purposes of assessment
- Characteristics of quality assessment
- Interrelationships between planning, teaching, learning, assessment and reporting
- Recognising, in practice, the scope and limitations of assessment, the benefits and the shortcomings
- Critical analysis of practice against principles of good assessment
- Familiarity with the actual requirements of schools and the location of those requirements (NAGs)
- Capability to make valid interpretations of requirements (NAGs).

Schools are often being told that “data” is the great panacea for guiding improvement. All data derived from measurement tests and assessments have a degree of error. Data is imperfect! We are well advised to exercise caution in interpreting data too literally.

Assessment is an *INEXACT* matter and can never be an exact one. Although we cannot aspire to exactness, there is every reason for not adding to inexactness.

Harlen, W. (ed). 1994. Enhancing Quality in Assessment. London: British Educational Research Association.

We should not expect to be able to measure pupils' abilities with the same confidence as we can measure their head size.

A definition of data for schools and teachers

Information sampled from multiple observations, examples of work, surveys, tests, etc. which give estimates of achievement in relation to learning goals or objectives.

Definition developed by a team of teachers
Assessment Workshop, Auckland, 2004.

Multiple Measures *a standard for quality*

The most helpful single principle in all testing is that test scores are data on which to base further study. They must be coordinated with background facts, and they must be verified by constant comparison with other available data.

Cronbach, L.J. 1970. *Essentials of psychological testing* (3rd ed.). (p. 381).

Multiple Measures

It is important for test users to realise that **no test score is perfectly accurate.**

Tests are primarily **samples** of behaviour and thus a single test of the kind commonly used in our schools can examine only a fraction of the knowledge and understanding, skills and competencies of a student.

Although the score on a test is determined principally by the ability and knowledge of the student who takes it, the score is also affected by the lack of precision inherent in all mental measurement.

A short test of, say, 10 questions tends to be unreliable because of its very limited sampling of a student's knowledge and abilities. Short tests are also affected by other factors such as chance variation in a student's motivation, concentration and mental set.

It is important, therefore, that test scores be regarded merely as estimates of a student's ability which are subject to errors of measurement. The result derived for an individual student from a single group test of achievement should **not** be regarded as final or sacrosanct.

Reid, 1993 (Progressive Achievement Test Manual).

No one test, measure, or assessment on its own is a sufficient judge of a student's progress and achievement.

**The Education Review Office
Ministerial Briefing
Literacy and Numeracy
November 2008**

New Zealand has a range of sophisticated world-standard assessment tools that have the potential to provide rich information to teachers, students and boards of trustees about students' progress. ERO expects that schools will use a range of assessment practices to gather information on the progress and achievement of students. p.1

The New Zealand **Curriculum**



Qualities of good assessment

The quality of assessment can be judged according to reliability, validity, manageability, and fairness.

Test developers' definitions

RELIABILITY

The degree to which a test consistently measures whatever it sets out to measure. Results from “form A are consistent with results from form B”.

VALIDITY

The degree to which a test measures what it is intended to measure, e.g. intelligence, aptitude, height.

The test developer's definitions of RELIABILITY and VALIDITY do not readily equate with classroom assessment purposes.

The test developer's intention is to devise measurement tools that differentiate students and/or show their levels of achievement.

The test developer's approaches are not necessarily the same as the teacher's. Therefore definitions of reliability and validity are needed that fit with nature and purposes of classroom teaching and assessment.

Assessment for Summing Up and Reporting

QUALITY INDICATORS

Reliability <i>strength</i>	Validity <i>strength</i>	Manageability <i>integrity</i>	Fairness <i>integrity</i>
Consistency of judgments	Interpretation & use (fullness) of information	Sensible use of time and resources	Alignment with goals



Dependability

The extent to which reliability is optimised while ensuring validity.

Assessment Systems for the Future (2006) www.assessment-reform-group.org

STRENGTH OF RELIABILITY

is reached through repeated observations and measures.

STRENGTH OF VALIDITY

is decided according to value of assessment information for teaching and learning (usefulness and impact).

While a single, one-off test might have good test design reliability, it does not follow that the results will necessarily give a reliable account of what a student knows and can do. The test is always only a sample, and the measure is always an approximation.



Harlen, W, 2007.
Assessment of
Learning. London,
Sage.

Reliability of teacher assessment

Threats to reliability need to be minimized. To increase the reliability of teachers' judgments, some internal school procedures for teachers to moderate each others judgments are necessary.

When it comes to **reliability**, assessment based on teachers' judgments is perceived as inferior to more controlled and formal methods, even though the latter are considerably less reliable than they are generally assumed to be. p.68

The New Zealand **Curriculum**



Assessment *Requirements*

The New Zealand Curriculum does not set out assessment and reporting requirements. Rather, it sets out expectations for learning.

The requirements for assessment and reporting are given in the **National Administration Guidelines**. These are stated in ways that allow schools interpret the requirements in ways that fit with their priorities and intentions. Importantly, these interpretations need to be shown to be consistent with the general intent of the National Administration Guidelines.

It is foreseen that the National Administration Guidelines will be changed by the Minister of Education and the Ministry of Education. Any changes will need to be consistent with the principles of a self-managing school system if those principles are to be upheld.

The National Administration Guidelines set the following requirements for assessment

Each Board of Trustees, through the principal and staff, is also required:

- To gather information that is **sufficiently comprehensive** to enable evaluation of student **progress** and **achievement**;
- To identify students and groups of students who are **not achieving**, who are **at risk of not achieving**, and who **have special needs**, and to identify **aspects of the curriculum** which require particular attention;
- In **consultation** with the school's **Maori community**, to develop and make known **plans and targets** for improving the achievement of Maori students.

These requirements give schools considerable responsibility and scope for deciding *what* they will assess, *how* they will assess, and the approaches and “tools” they will choose to use.

The New Zealand **Curriculum**



Teacher Assessment

Teacher Assessment

Teacher assessment is **essentially an informal activity**: the teacher may pose questions, observe activities, evaluate pupils' work in a planned and systematic or ad hoc way. **The information which the teacher thus obtains may be partial or fragmentary; it will not at the time allow the teacher to make a firm evaluation of the pupils' competence in reading, for example, or understanding of a mathematical process. But repeated assessment of this sort, over a period of time, and in a range of contexts will allow the teacher to build up a solid and broadly-based understanding of the pupil's attainment.**

Gipps, 1994.

Harlen, W. (ed). 1994. Enhancing Quality in Assessment. London: British Educational Research Association.

Teacher Assessment

Teacher assessment is **essentially an informal activity**: the teacher may pose questions, observe

activities, use a system of teacher fragments to make in real mathematics

This premise is only partly satisfactory. The quality of ALL teacher assessment requires constant reference to pre-determined learning goals *AND* knowledge of what student work should be like relative to particular stages of learning or progress in relation to those goals.

in the teacher's presence
of

this sort, over a period of time, and in a range of contexts *will* allow the teacher to build up a solid and broadly based understanding of pupil attainment.

Gipps, 1994.

Generalisability of Teacher Assessment

If the assessment has sampled broadly across the domain (learning area) *and* in depth, within it, then the assessment is likely to be generalisable. ... An external test, on the other hand, will provide more limited information based as it is on a one-off occasion covering a limited sample of tasks.

Gipps, 1994.

Harlen, W. (ed). 1994. Enhancing Quality in Assessment. London: British Educational Research Association.



Harlen, W, 2007. Assessment of Learning. London, Sage.

Using assessment for learning has been included in the text of various ‘new strategies’ introduced by the Department for Education and Skills in England. While this is to be welcomed there is some evidence that the interpretation and implementation in practice tend to be **mechanistic and lacking in signalling the essential shift in pedagogy** that using assessment for learning requires. p.3

Does this observation apply to practice in your school?

FORMATIVE ASSESSMENT

Date: February 2008 Learning Area: Mathematics - Statistics

Expert: (E) High level of mastery
Practitioner: (P) Understanding with mastery
Apprentice: (A) Understands but needs teacher assistance.
Novice: (N) Limited understanding

Assessment Type: (Highlight)

Performance	Observation	Interview	Pen/Paper	Peer	Self	Journal	Checklist
	Conference		Assessment task		Group Assessment		

The student is able to:

- ⇒ Conduct a statistical enquiry by asking questions to gather data.
- ⇒ Sort and display data using a tally chart, pictograph, bar graph, stem and leaf graph etc.
- ⇒ Compare data and make statements about a particular graph.

GROUP: Whole Class

NAME	1	2	3	4	Anecdotal Notes:	Pre-Test	Post-Test
	✓	✓	✓	P	60%	7	26
	✓	✓	✓	P	74% Recap Pictographs	9	32
	✓	✓	✓	E	94%	25	40.5
	✓	✓	✓	P+	82%	12	36
	✓	✓	✓	E	90%	16	38.5
	✓	✓	✓	A	42% No idea how to record data in a stem leaf graph or interpret	4	18
	✓	✓	✓	P	78%	13	33.5
	✓	✓	✓	P	80%	9	34.5
	✓	✓	✓	E	95%	16	41
	✓	✓	✓	E	95%	18	41
	✓	✓	✓	P	67%	ab	29
	✓	✓	✓	P+	83%	11	35.5
	✓	✓	✓	P	77%	10	33
	✓	✓	✓	E	95%	12	41
	✓	✓	✓	P	71%	15	30.5
	✓	✓	✓	P	69%	2	29.5
	✓	✓	✓	P+	85%	17	36.5
	✓	✓	✓	P	77%	11	33
	✓	✓	✓	P	73% Recap Stem & Leaf	13	31.5
	✓	✓	✓	A/P	58% Recap Stem & Leaf	8	25
	✓	✓	✓	A	45% Attempted SL graph but did not attempt to interpret it.	3	19.5
	✓	✓	✓	A	49% Didn't attempt SL graph questions & last pictograph	9	21
	✓	✓	✓	P+	84%	11	36
	✓	✓	✓	P	74%	11	32
	✓	✓	✓	P+	83%	16	35.5
	N/A	N/A	N/A	N/A	N/A	10	N/A
	N/A	N/A	N/A	N/A	N/A	16	N/A
Connor	✓	✓	✓	A	51% Recap Stem & Leaf	N/A	22

Is this example substantially "formative" assessment, or substantially "summative" assessment? Are these distinctions helpful or necessary?



Harlen, W, 2007.
Assessment of
Learning. London,
Sage.

Learning about how to learn and the ability to reflect on the adequacy of what one knows is the key to taking steps towards further learning. Research shows that the ability to take effective action results from students being helped to:

- see how to improve their work, by feedback that is non-judgmental
- try to explain things rather than just describe them
- take some responsibility for assessing their own work, finding the errors in their own or a partner's work
- talk about and justify their reasoning
- understand the goals and the quality of work they should be aiming for.

Is this the kind of practice at work in your school?



Different approaches to summative assessment by teachers vary most significantly in the extent to which, on the one hand, there is **specification of the sources of evidence** and, on the other hand, **specification of the criteria** by which the evidence from these tasks is turned into levels, grades, scores, or rankings.

1. **High level of specification of both task and criteria**
e.g. portfolio samples.
2. **High level of specification of task, low level of specification of criteria**
e.g. portfolio - sample items specified, minimal guidance on marking such as 1 - 5 scale.
3. **Low level of specification of both task and criteria**
e.g. teacher given freedom to select evidence, criteria are general.
4. **Low level of specification of task and high level of specification of criteria**
e.g. freedom to select evidence combined with generic criteria (which can help guide the selection of work assessed).

Which of these four approaches is likely to provide greatest dependability of teacher judgment?



Harlen, W, 2007.
Assessment of
Learning. London,
Sage.

Greater dependability of teacher judgments is attained when there are detailed, but generic, criteria that allow evidence to be gathered from the full range of classroom work.

The criteria guide the selection of evidence without prescribing it!

When criteria are well specified, teachers are able to make reliable judgments.

Studies give no evidence that assessment by teachers is more or less reliable in certain subjects rather than others.

pp.72-73



Harlen, W,
2007.
Assessment of
Learning.
London, Sage

Summary of conditions affecting dependability

- Detailed criteria describing levels of progress in various aspects of achievement enable teachers to assess students reliably on the basis of regular classroom work.
- It is important for teachers to follow agreed procedures if teachers' assessment is to be sufficiently dependable for summative purposes.
- The training of teachers to improve the reliability of their assessment is more effective when it involves teachers as far as possible in the process of identifying criteria so as to develop ownership of them and understanding of the language used.
- Training should also focus on the sources of potential bias.
- Dependable assessment needs protected time for teachers to meet and to take advantage of the support that others can give.
- Moderation through professional collaboration benefits teaching and learning as well as assessment.

pp.75-76

The New Zealand **Curriculum**



Student Self-Assessment

Student self assessment



Harlen, W, 2007.
Assessment of
Learning. London,
Sage.

Giving students some role in assessing their learning is central to promoting learning autonomy, which would seem to be the most secure foundation for lifelong learning!

p.126

Key Premise for Self-Assessment

A key premise is that for students to be able to improve, they must develop the capacity to monitor the quality of their own competencies during actual performance. This in turn requires that students possess an appreciation of what a competency means, that they have been helped to have the evaluative skill necessary for them to compare with some objectivity the quality of what they are doing in relation to particular criteria, and that they develop a store of tactics which can be drawn upon to improve their own performance.

Sadler, R. 1989. Formative assessment and design of instructional systems, *Instructional Science*. 18: 119-144. (modified)

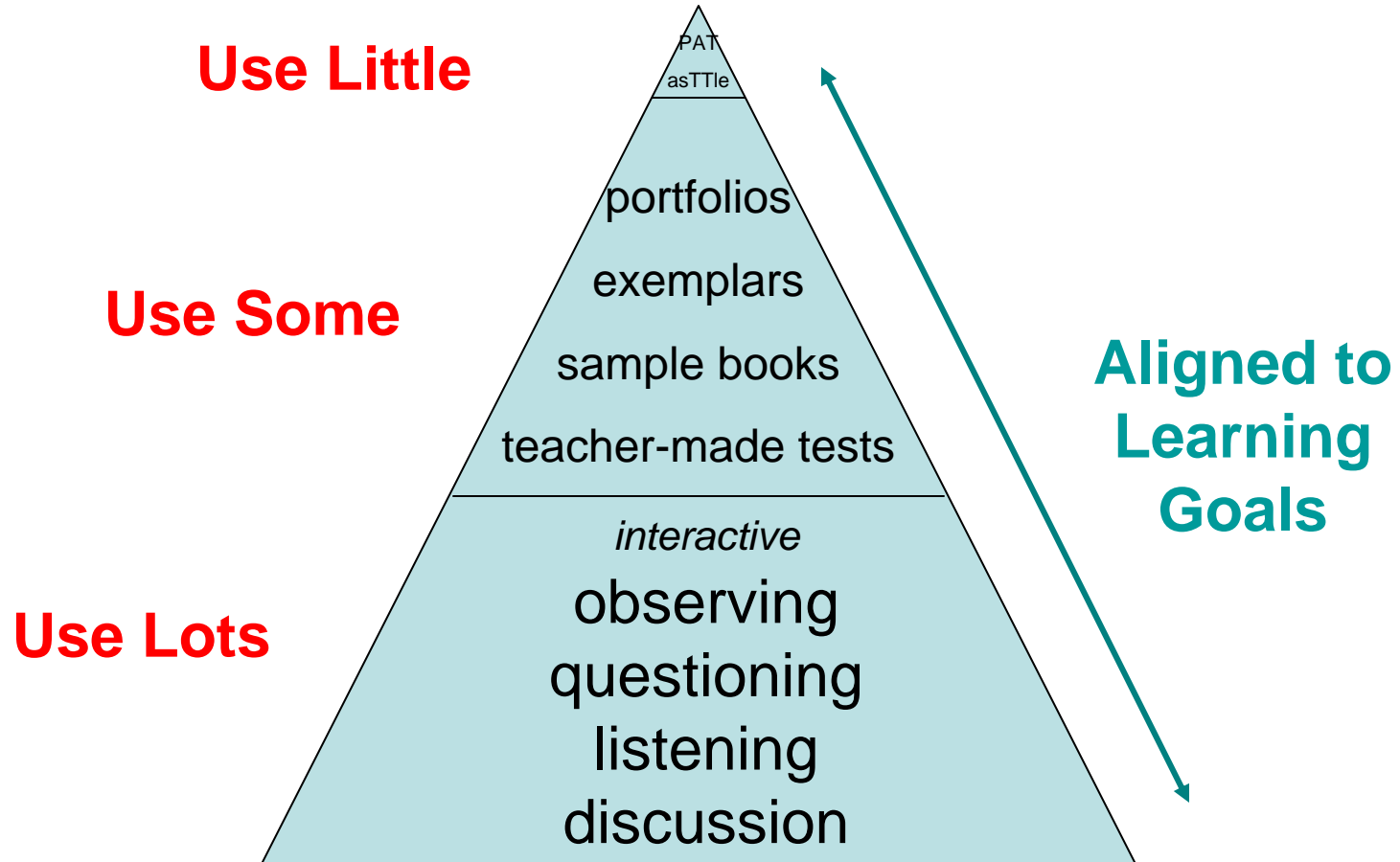
The New Zealand **Curriculum**



Healthy Assessment

The Healthy Practice Pyramid

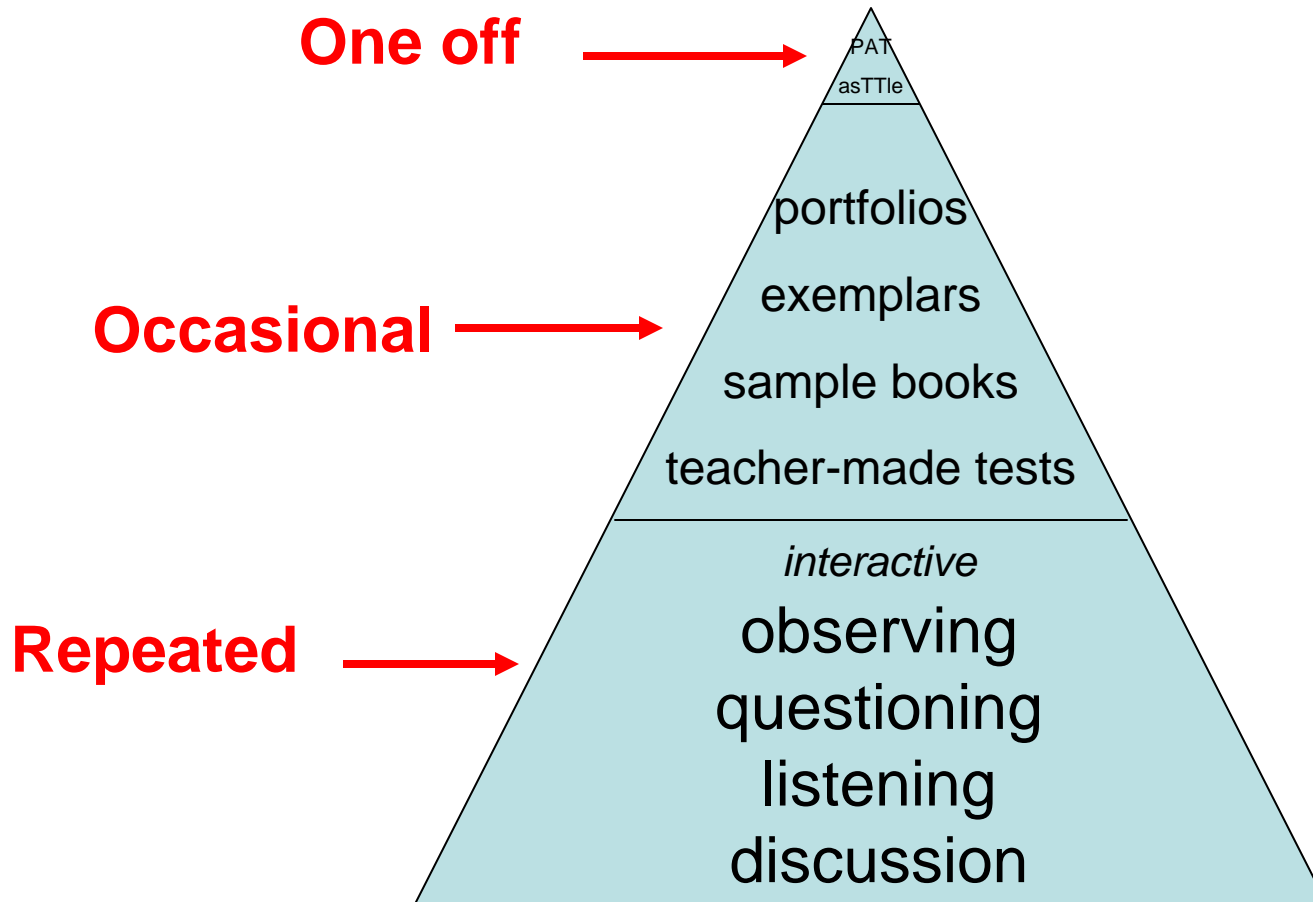
Classroom Assessment Information Sources



For **STRENGTH** of information, use multiple *samplings* from multiple sources.

The Healthy Practice Pyramid

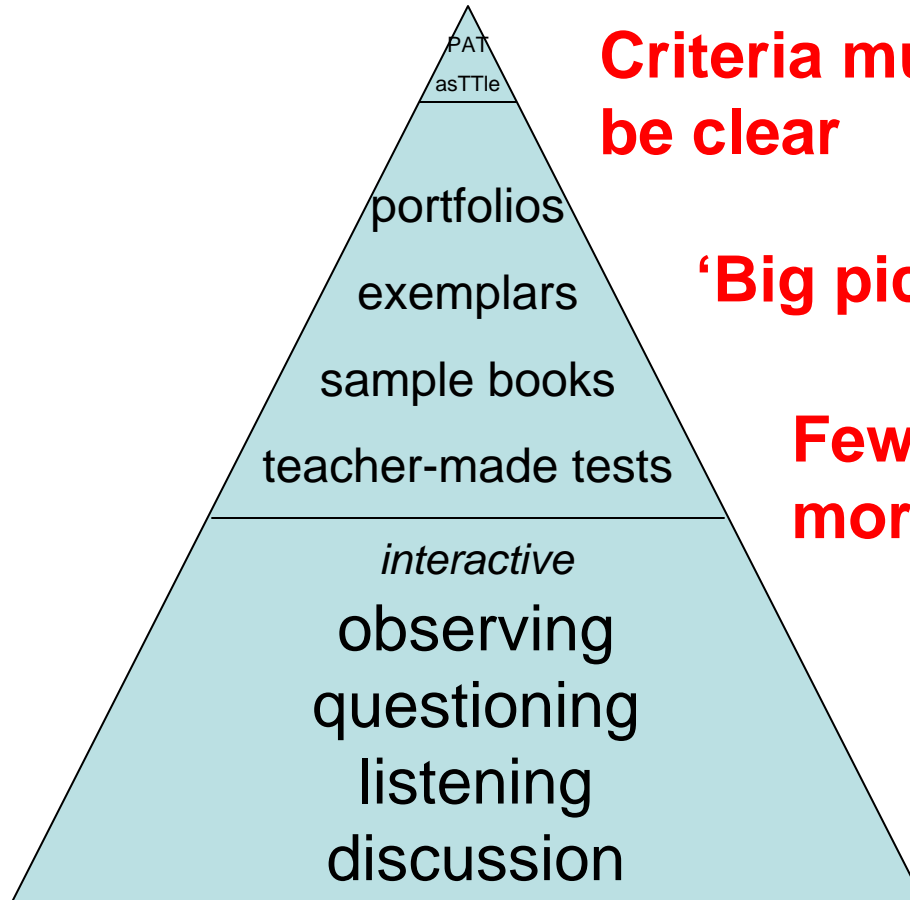
Classroom Assessment Information Sources



For **STRENGTH** of information, use multiple *samplings* from multiple sources.

The Healthy Practice Pyramid

Classroom Assessment Information Sources



**Criteria must
be clear**

'Big picture' criteria

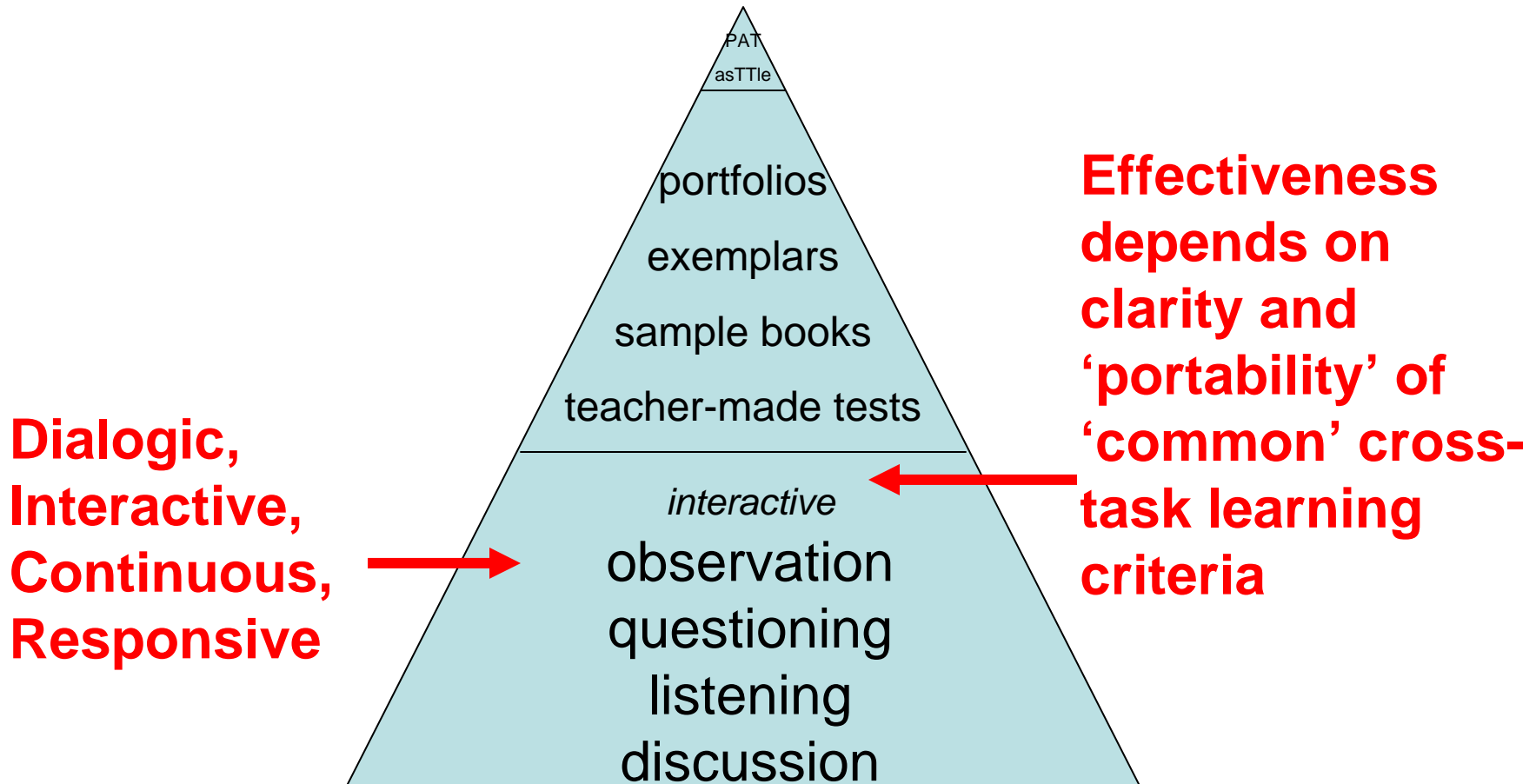
**Fewer rather than
more**

**Criteria apply
across
learning tasks**

For **STRENGTH** of information, use multiple *samplings*
from multiple sources.

The Healthy Practice Pyramid

Classroom Assessment Information Sources



For **STRENGTH** of information, use multiple *samplings*
from multiple sources.