Performances

ASSESSMENT RESOURCE KIT

'Doing' Outcomes
Planning Observations
Judging & Recording
Estimating Achievement
Reporting

Informal Observations
Assessment Events
When Comparability Matters
Designing Assessments

ISBN 0-86431-207-5

ARK

Margaret Forster & Geoff Masters

Contents

First published 1996
by The Australian Council for
Educational Research Ltd.
19 Prospect Hill Road, Camberwell, Melbourne,
Victoria 3124, Australia.

Reprinted 1998

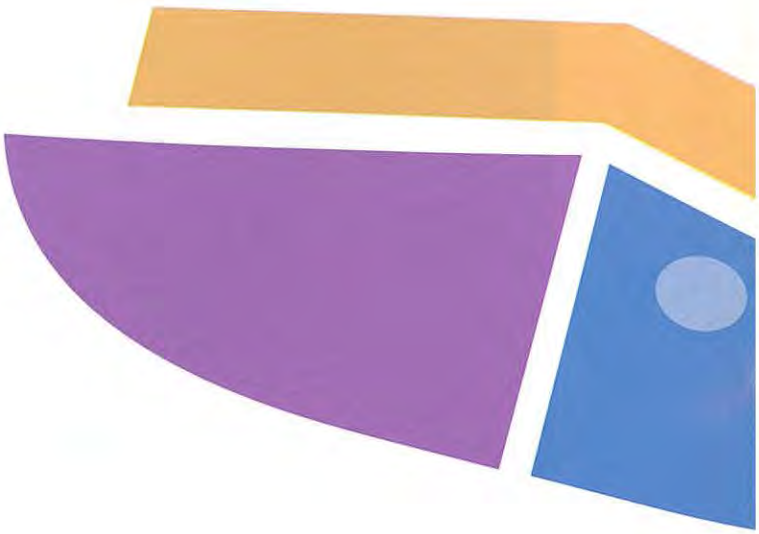
This project was supported by a Grant under the Curriculum Development Projects Program from the Commonwealth Department of Employment, Education and Training and Youth Affairs. The views expressed here are those of the authors and do not necessarily represent the views of the Department.

ISBN 0 86431 207 5

Edited by Eve Recht
Designed by Benjamin Forster

Printed by Allanby Press Printers Pty. Ltd.

Photographs courtesy of Melbourne High School, South Yarra; Methodist Ladies College, Kew; Village School, North Croydon; Knox Grammar, Wantirna South; and Lisa Norwick, Seattle.



what is performance assessment?

The assessment of some kinds of learning requires the observation of student performances. Examples include playing a musical instrument, dancing, using scientific apparatus, operating equipment, performing gymnastics routines, diving, problem-solving in a group, applying ball-handling skills, and participating in small group discussions.

'Performance assessment' is the assessment of students *as they engage in an activity.* It is the on-the-spot evaluation of performance, behaviour or interaction. There is no concrete product that can be judged at a later stage, unless perhaps it is an audio- or video-taped record of the original performance.

'doing' outcomes

Curriculum documents include many learning outcomes which describe processes, activities or performances. These outcomes can be described as 'doing' outcomes. To assess these outcomes it is necessary to observe students in action. This kind of assessment is especially important in some learning areas, such as the Arts and Physical Education, and in some learning strands such as Speaking, in English.

Other definitions

Performance assessment is sometimes defined more loosely as any assessment requiring students to construct or generate a response. Short answer questions and essays, for example, are sometimes called 'performance assessments'. This is not what we mean.

The term performance assessment is also sometimes used to refer to 'authentic' assessments of real world, on-the-job performances. This is not what we mean either.

Performance assessment, as defined here, is assessment which relies on the observation and judgement of activities as they occur.

Using Tools and Equipment

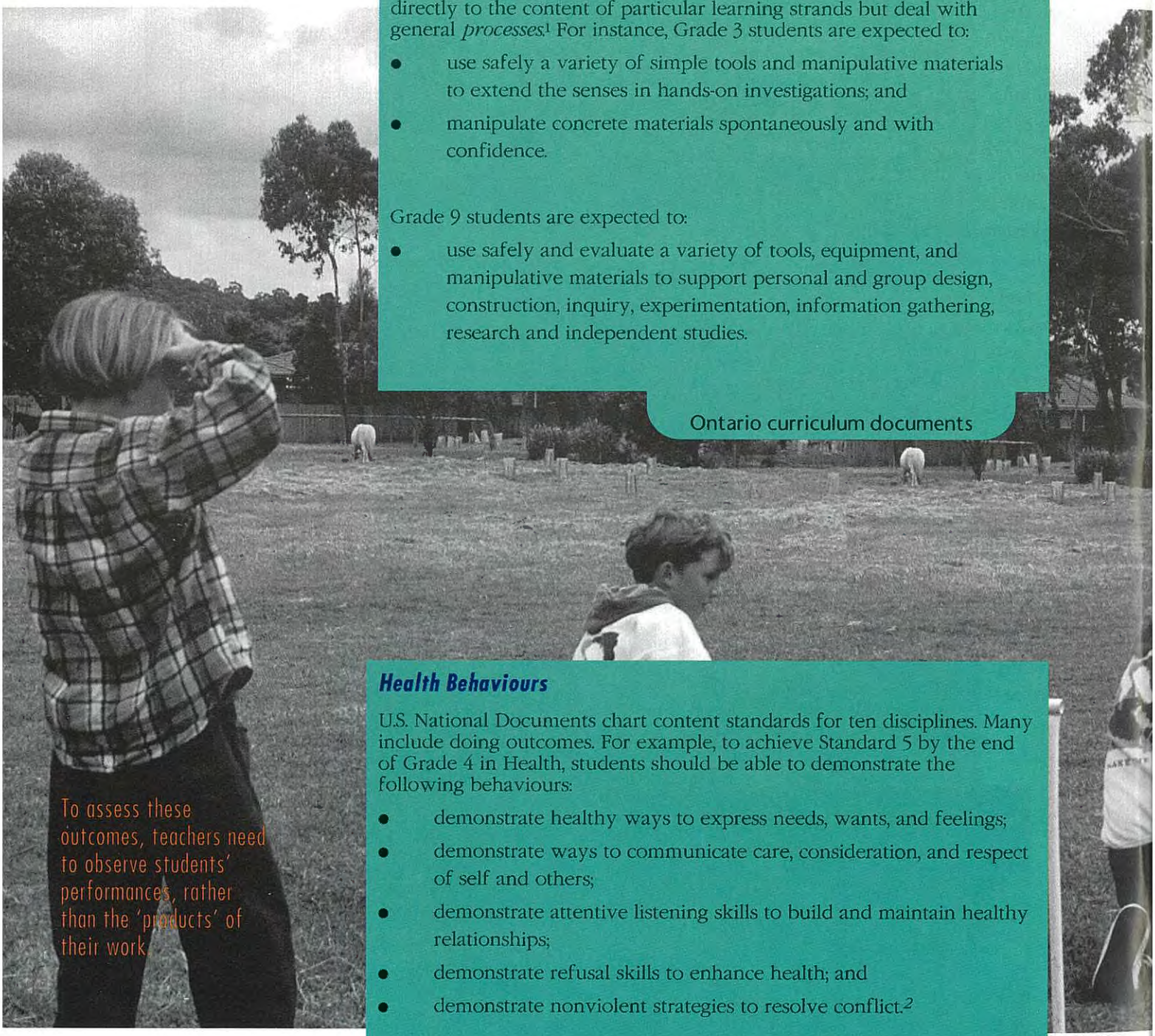
Some outcomes in the Ontario Common Curriculum do not relate directly to the content of particular learning strands but deal with general *processes*.[1] For instance, Grade 3 students are expected to:

- use safely a variety of simple tools and manipulative materials to extend the senses in hands-on investigations; and
- manipulate concrete materials spontaneously and with confidence.

Grade 9 students are expected to:

- use safely and evaluate a variety of tools, equipment, and manipulative materials to support personal and group design, construction, inquiry, experimentation, information gathering, research and independent studies.

Ontario curriculum documents



To assess these outcomes, teachers need to observe students' performances, rather than the 'products' of their work.

Health Behaviours

U.S. National Documents chart content standards for ten disciplines. Many include doing outcomes. For example, to achieve Standard 5 by the end of Grade 4 in Health, students should be able to demonstrate the following behaviours:

- demonstrate healthy ways to express needs, wants, and feelings;
- demonstrate ways to communicate care, consideration, and respect of self and others;
- demonstrate attentive listening skills to build and maintain healthy relationships;
- demonstrate refusal skills to enhance health; and
- demonstrate nonviolent strategies to resolve conflict.[2]

US National Standards

Movement

Curriculum Profiles for Australian Schools detail outcomes for eight key learning areas. Many of these outcomes are 'doing' outcomes.

Health and Physical Education

'Demonstrates coordinated actions of the body by performing and modifying movement sequences.'

Evident when students, for example:

- adapt the speed, force, flow, direction and height of movement to improve performance; and
- vary running movement patterns to cater for sprinting, hurdling and distance running.[3]

Role Play and Working in a Group

Studies of Society and Environment

'Expresses a personal view of the meaning of data.'

Evident when students, for example:

- role-play a character from a text to show the character's responses in similar situations.

'Identifies causes of conflict and ineffective group work and negotiates solutions.'

Evident when students, for example:

- resolve conflicts and problems in a group without resorting to a person in authority; and
- put personal feelings aside to complete tasks (work with peers other than friends to share skills and knowledge).[4]

Dance

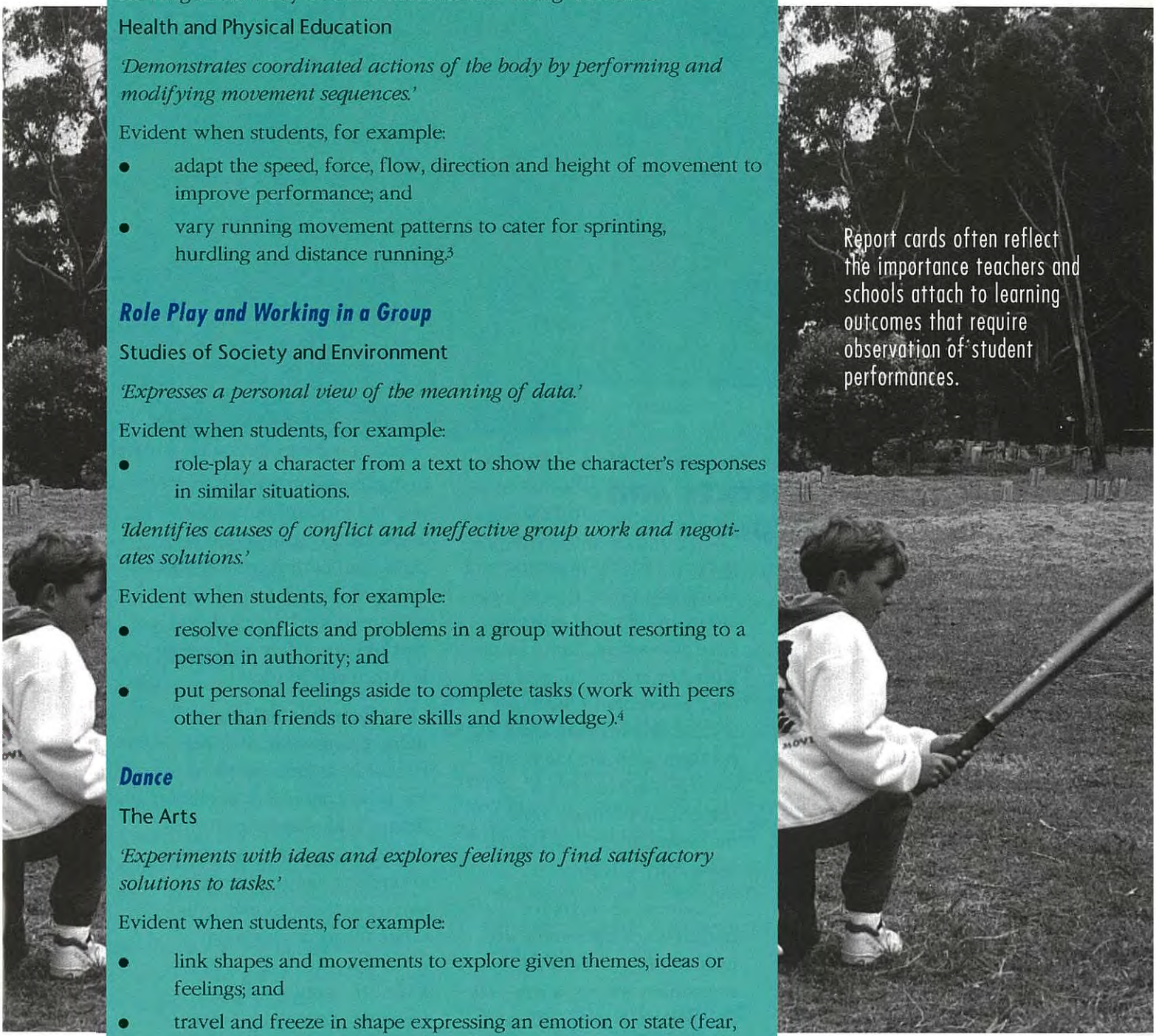
The Arts

'Experiments with ideas and explores feelings to find satisfactory solutions to tasks.'

Evident when students, for example:

- link shapes and movements to explore given themes, ideas or feelings; and
- travel and freeze in shape expressing an emotion or state (fear, hunger, joy) and link several of the shapes to make a dance sequence.[5]

Curriculum Profiles for Australian Schools



Report cards often reflect the importance teachers and schools attach to learning outcomes that require observation of student performances.

Using Instruments

Mathematics

'Measures and makes things, using a range of graduated scales and strategies for making measurements that are more accurate than the available equipment allows.'

Evident when students, for example:

- use instruments correctly and accurately (place a measuring jug on a level surface and read with markings at eye level to reduce reading error).[6]

Musical Performance

The Arts

'Rehearses, presents and promotes musical works in ways appropriate for particular audiences.'

Evident when students, for example:

- rehearse and perform musical works as a member of an ensemble demonstrating the ability to maintain an independent part and to blend sensitively and balance within the ensemble while performing as either conductor, leader or group performer.[7]

Curriculum Profiles for Australian Schools

contexts and purposes

There are many contexts and purposes for the assessment of 'doing' outcomes; that is, there are many *kinds* of performance assessment. For example, a teacher engages in performance assessment when, in discussing a mathematics problem with a student, she observes and records a response reflecting a new understanding: 'Now I understand why ¼ is 25%!'

Classroom contexts for performance assessment also include more structured assessment 'events', a role-play in Japanese, a small group drama performance, for example, or a Physical Education assessment of ball-handling skills. And performance assessments of a more formal and structured kind are made as part of high-stakes assessments in dance and music.

The way in which performance assessments are planned, the way in which student performances are judged, and the way in which the resulting evidence is used to infer students' locations on a progress map, depend on the context and purpose of assessment.

It is useful to think about these different kinds of performance assessment along a 'purpose' continuum, as shown below, with classroom formative and diagnostic assessment at one end and high-stakes summative assessment at the other. Informal classroom observations are towards the left of the continuum; teachers' planned classroom

assessment events – a teacher-student conference or a group observation in science, for example – are in the centre of the continuum, and externally set high-stakes assessments are to the right.

The way in which these different kinds of performance assessments are planned, the way in which student performances are judged, and the way in which the resulting evidence is used to infer students' locations on a progress map, depend on the context and purpose of the assessment.

planning performance assessments

The usefulness of the evidence provided by performance assessments for inferring a student's level of achievement depends on

- relevance and coverage (a validity issue); and
- amount of evidence (a reliability issue).

relevance and coverage

'Relevance' describes the degree to which the evidence addresses knowledge, skills and understandings, or outcomes, of the learning area.

Classroom teachers are likely to be interested in ensuring that observations of student performance focus on instructional goals and that there is a match between planned performance tasks and the outcomes to be assessed.

'Coverage' describes the degree to which the evidence samples the range of outcomes in a learning area. Teachers usually observe a variety of student performances in an attempt to collect evidence about a range of outcomes.

Teachers are likely to be interested also in developing ways to efficiently record anecdotal information about student performance to provide systematic evidence to guide instruction and feedback to students and parents.

amount of evidence

Although assessments of student achievement are based on observations of their performances on specific tasks, the purpose and interest always is in the knowledge, skills and understandings required to perform those tasks. The intention is to infer a student's achievement in an area of learning from their performances on specific tasks, that is, to generalise from a limited performance sample.

This observation raises the question of the amount of evidence (number of tasks) required to provide a sufficiently accurate estimate of a student's general level of achievement in an area of learning. This question is particularly important in high-stakes contexts in which student performances influence admission, scholarship offers and the award of certificates.

judging performances

When performances are judged in order to infer a student's level of achievement, two features of the assessment need to be considered:

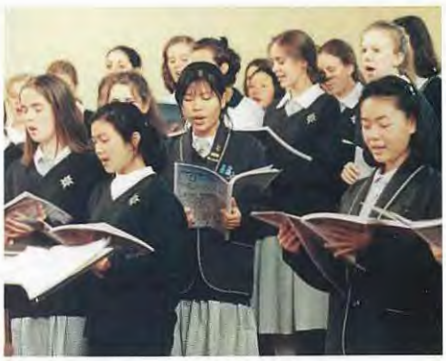
- the method of judgement; and
- the comparability of judgements (inter-rater reliability).

method of judgement

Performances are judged either holistically or analytically. Teachers make an *holistic* judgement of a student's performance when they give a single rating based on their overall impression of that performance.

Performance assessments provide evidence of student achievement within a learning area. The usefulness of this evidence for inferring a student's level of achievement depends on:

- relevance and coverage (a validity issue); and
- amount of evidence (a reliability issue).



When performances are judged in order to infer a student's level of achievement, two features of the assessment need to be considered:

- the method of judgement; and
- the comparability of judgements (inter-rater reliability).

Teachers make an *analytic* judgement of a student's performance when they rate different aspects of the performance. A dance performance, for example, might be rated for content and performance features – ideas, organisation, pace of delivery, volume, facial expression, and audience engagement. Teachers may report these ratings separately or use them to make a final on-balance assessment of the performance.

comparability of judgements

A fundamental feature of performance assessment is its reliance on judgement. Two people viewing the same performance can judge that performance differently. To ensure that assessments are fair – that is, that the rating of a student's performance does not depend on who assesses it – it is important to minimise differences among assessors: in other words, to ensure inter-marker reliability.

In the classroom context, whether another teacher would make the same judgement of a student's performance using the same marking guide is not of great concern. In high-stakes contexts, however, inter-marker reliability is crucial.

Usually, the greater the requirement for comparability of performance assessments, the more precisely the assessment criteria are specified. In a high-stakes situation such as a final year of school assessment, the criteria for assessment may be tightly specified. Markers may be trained and the marking process carefully monitored to ensure a high level of inter-marker agreement.

estimating and reporting locations on a progress map

In developmental assessment (see *ARK Developmental Assessment*) teachers monitor student progress against a preconstructed map of developing skills, knowledge and understandings. Teachers make observations of students' performances in contexts that are relevant to the learning area and to the outcomes described on that map. These observations are the 'evidence' used to estimate students' levels of attainment. The greater the number of observations, the more dependable the conclusion about each student's level of attainment.

Developmental assessment requires an *on-balance* decision (inference) about a student's location on a progress map based on the available evidence. The way in which this inference is made is usually determined by the purpose of the performance assessment. The higher the stakes, the greater the requirement for comparability, and the more tightly the 'inference process' is likely to be specified.

subjective estimates

Performance assessments are one source of evidence from a variety of methods teachers use to observe, record, and collect evidence. A teacher of Health and Physical Education might make an inference about a student's achievement in 'Human Movement', for example, using a combination of evidence from anecdotal records of student

performances in class, paper and pencil assessments, and observations on performance tasks. In some instances a student's location on a progress map might be estimated entirely from their performances on assigned tasks.

When the estimate of a student's level of achievement is made subjectively, there may be only a loose connection between the available evidence (observations and judgements of performance) and the resulting estimate of the student's location on a progress map.

objective estimates

In high-stakes situations, where high levels of inter-marker comparability are desirable, the way in which judgements are combined to estimate a student's location on a progress map may be tightly prescribed to ensure that the estimate is made in the same way by different assessors (i.e. objectively). The inference may be made numerically on the basis of a marker's pattern of judgements across carefully defined criteria, for example.

The following three articles explore different kinds of performance assessment:

'Informal Observations' describes the day-to-day classroom assessment of 'doing' outcomes and two strategies for improving the quality of information:

- focusing observations; and
- recording observations efficiently.

'Performance Assessment Events' describes strategies for planning, judging and recording performances. Examples include:

- self-rating performance on a collaborative problem solving task;
- self-assessment of participation in a group task;
- analytic rating of music performance;
- holistic assessment of speaking; and
- analytic rating of drama performance.

'When Comparability Matters' considers the use of performance assessments in contexts where the comparability of assessments is particularly important. Issues in these contexts include:

- curriculum match;
- fairness;
- generalisability;
- standardisation; and
- inter-rater reliability.

For discussion:

'Just because an assessment asks students to perform an interesting or complex activity does not make it a good assessment. Good assessment reliably measures something beyond the specific tasks that students are asked to complete. The results of good assessment identify what students can do in a broad knowledge or skill domain. The skills that students exhibit in the assessment situation should transfer to other situations and other problems.'[8]

1 Ministry of Education and Training (1993) The Ontario Common Curriculum, Grades 1-9.

2 *Education Week* 12 April 1995 Washington, Special Report 'Struggling for Standards', p. 59.

3 Curriculum Corporation (1994) *Health and Physical Education - A Curriculum Profile for Australian Schools*, Carlton: Curriculum Corporation, p. 86.

4 Curriculum Corporation (1994) *Studies of Society and Environment - A Curriculum Profile for Australian Schools*, Carlton: Curriculum Corporation, pp. 37 & 85.

5 Curriculum Corporation (1994) *The Arts - A Curriculum Profile for Australian Schools*, Carlton: Curriculum Corporation, p. 70.

6 Curriculum Corporation (1994) *Mathematics - A Curriculum Profile for Australian Schools*, Carlton: Curriculum Corporation, p. 90.

7 Curriculum Corporation (1994) *The Arts - A Curriculum Profile for Australian Schools*, Carlton: Curriculum Corporation, p. 114.

8 Herman, J., Aschbacher, P. & Winters, L. (1992) The National Center for Research on Evaluation, Standards and Student Testing (CRESST). *A Practical Guide to Alternative Assessment*, Alexandria, VA: Association for Supervision and Curriculum Development, p. 9. Copyright © by The Regents of the University of California and supported under the Office of Educational Research and Improvement (OERI), US Department of Education.

informal classroom observations

Performance assessment is the assessment of students *engaged in an activity* – diving, using scientific apparatus, discussing a story, throwing a javelin. There are many contexts and purposes for performance assessment, from informal, ongoing classroom assessment to formal, high-stakes assessment. This article looks at informal performance assessment: the assessments teachers make in the classroom *as they are teaching* rather than in planned assessment 'events'.

Observations of student behaviour are necessary for some kinds of outcomes which cannot readily be assessed by looking at the products of student work. Assessing a student's ability to work effectively in a group, or skill in using a particular technique in science, for example, requires observations of student performance. Other examples include: using mathematical instruments, demonstrating running techniques, and holding dance positions.

Teachers have many opportunities to observe student performances. They watch students participating in class activities, contributing to class discussions and interacting with other students. Much of the information gained in this way is used immediately to inform practice, to respond to students' special needs, and to add to the overall picture of a student's progress as a learner. But the process is often haphazard, relying on chance observations.

The usefulness and quality of information obtained from day-to-day observations of student performances can be improved by:

- focusing observations on important learning outcomes; and
- recording observations systematically.

focusing observations

It is not practical to record comments on every pupil, every day, about everything said or done. Observations must be focused on *significant events as they relate to relevant learning outcomes*. 'Significant' events might be:

- unexpected student behaviour;
- a clear illustration of understanding or lack of understanding; or
- clear evidence concerning a particular learning outcome.

spotlighting

One technique for focusing observations on important learning outcomes is 'spotlighting'. Spotlighting is the process of watching for just a few aspects of student behaviour. Examples of spotlighting oral language and psychomotor behaviour are shown on page 10.

Teachers sometimes choose to spotlight outcomes from different learning areas in the same activity. A teacher might target Working Scientifically outcomes *and* Speaking and Listening outcomes in a science task that requires students to work in a small group using equipment to investigate a problem, for example.

The usefulness and quality of information obtained from day-to-day observations of student performances, as a basis for monitoring student progress, can be improved by:

- focusing observations on important learning outcomes; and
- recording observations systematically.

Observations must be focused on *significant events as they relate to relevant learning outcomes*:

- unexpected student behaviour;
- a clear illustration of understanding or lack of understanding;
- clear evidence concerning a particular learning outcome.

To obtain the most information possible from day-to-day observations teachers need to develop a recording system which is:

- organised;
- easy to use; and
- time efficient.

It also needs to:

- give ready access to information; and
- allow growth over time to be seen.

Focusing observations — Spotlighting

Terry Spottiswood has developed a checklist for observing oral language which focuses on just four student behaviours. She watches for evidence of these four behaviours in small group interactions.[1]

Name	Indicates disagreement in a positive manner	Supports statements of others	Asks questions to draw out information	Listens without interrupting
Georgio	✓		✓	✓
Dean			✓	
Sam	✓	✓	✓	✓
Wai-Ming	✓	✓	✓	

Focusing observations — Spotlighting psychomotor behaviours

Roy Killen suggests that a checklist for focusing observations on psychomotor behaviours include the sequence of steps students are expected to follow in classroom activities.[2]

Using a Beam Balance	Check
• Places object of unknown weight on the left side of the balance.	❑
• Estimates weight and places appropriate known weight on the right side of the balance.	❑
• Raises balance arm and checks for balance.	❑
• Notes whether weight was too large or too small.	❑
• Lowers balance.	❑
• Adds or subtracts known weights (as appropriate) and rechecks balance.	❑
• As balance is approached, adds and subtracts lesser weights.	❑
• Brings the scale closer to balance with the addition or subtraction of each weight (rather than trying to achieve balance in a hit-or-miss fashion).	❑
• When balance is achieved, adds the balance weights to get the correct weight of the object.	❑

When spotlighting behaviours, teachers often focus their observations at different stages of a process. For example, if students are learning a series of skills in isolation before combining them for a complex performance (e.g. a gymnastics floor routine) a teacher may spotlight performance of one or two of the isolated skills, as well as aspects of the final performance.

It also can be helpful to 'spotlight' particular students rather than attempt to observe all students every lesson. Teachers might choose to observe five students each day, for example. Working systematically through the class in this way helps alert teachers to 'invisible' students about whom little might otherwise be recorded.

focused questioning

Focused questioning is a technique that can be used to gather additional information while 'spotlighting' students. Teachers use focused questioning to gain insights into a student's conceptual grasp of a task: Why are they attempting to solve a problem in a particular way? How do their findings relate to their predictions and to other experiences they have had? What do they understand about the skills required to complete a complex series of movements?

recording observations

To obtain the most information possible from day-to-day observations, teachers need to develop a recording system which is:

- organised;
- easy to use; and
- time efficient.

It also needs to:

- give ready access to information; and
- allow growth over time to be seen.

There are many ways to record classroom observations including file cards, computer records, and observation grids and sheets. Teachers may need to try several methods, or a combination of methods, or invent their own, before they find the system that best suits them, their subject area, and their classroom.

file cards

Teachers have developed a variety of ways to use file cards to record informal performance assessments. Ann Beyer suggests the following:[3]

1 Keep cards in a folder. Use one card for each student for each learning area. Write observations directly on cards and date them. When cards are filled, staple extra cards on.

2 Keep stapled set of cards, for each student, in a folder. Different coloured cards may indicate different strands of a learning area.

3 Tape cards on to a single sheet and keep on a clipboard for easy use. Once a card is full, file it for future reference.

computer records

If teachers have access to a computer in the classroom they can enter observations directly on to the computer. Using 'cut and paste' editing functions they can then transfer their observations to files for individual students. These can be imported into report files at a later date.

annotated class lists

Annotated class lists can be used to record chance observations of student behaviours.

observation grids or sheets

Observations of student performances also can be recorded on class grids or sheets. These can be used for a particular period of time or a particular task. Some teachers paste the grids inside a folder for privacy. Others put them on the classroom wall so that students can see the outcomes being assessed.

Some teachers cut up the class sheet at the end of each week and transfer the individual grid entries to each student's record folder. Other teachers have refined the system for easier and more efficient recording. They write their observations on *'post-it'* notes in the grid squares (see opposite). The post-it notes then can be transferred easily to a record book.

Recording observations — Observation grids

Kevin Olsen and his colleagues describe one teacher's use of A3 sheets for observation grids.[4] At the end of each week the teacher dates the entries, cuts them up, and has students glue them on to a card in their folders. It is important that comments placed on a grid of this kind 'stand alone'; that is, that comments on a student mean something later when they are separated from the context of the surrounding comments. The observations below relate to mathematical understandings, but this system could be used for recording observations of students' performances in any learning area.

Rhys 20/5 – animals, natural objects patterns 1st, 2nd '... and the sheep is first and the chook is second ...' (N + WM – L1) 21/5 – time sequence in a day fairly accur. uses first and second to describe sequence – record orally and draw (M + WM – L1)	Xenia	Antonia	Kate
Chris	Con	Wayne 17/5 – patterns – loves colours. uses them as ones – 'first this one here, then the blue's next' – records accurately matching colours (N + WM, L1) 24/5 – daynight – ok – 'dinnertime's when mum comes in' – [illegible] (WM, L1)	Rose 17/5 – patterns – multiple combinations uses 1st only 'this one is first, then this' draws + numbers them 1, 2, ... (N, L1) 27/5 – chance and data – weather map – making predictions 'today it rained, 9 thigh – will tomorrow, today
... and the other 23 children			

Recording observations

Post-it notes

Bonnie Campbell Hill has refined the grid system for easier and more efficient recording using post-it notes:[5]

Teachers using her system photocopy a form with blocks the size of small post-it notes and space to write each student's name. Teachers write in students' names then run off several copies of the form. Blank post-it notes are placed in the space below each name. The teacher carries the form on a clip board and notes observations directly on to the post-it notes. These are dated. At regular intervals the post-it notes are transferred to each student's record sheet. These are kept together in a separate folder. (The same process can be used with sheets of mailing labels. These work better in that they stick more firmly. However, they can be moved only once.) The observations below range from general attitudes to mathematics, to writing behaviours. This system can be used to focus observations on behaviours in any learning area.

ANECDOTAL RECORDS for Cindy's Conference Group (teacher)

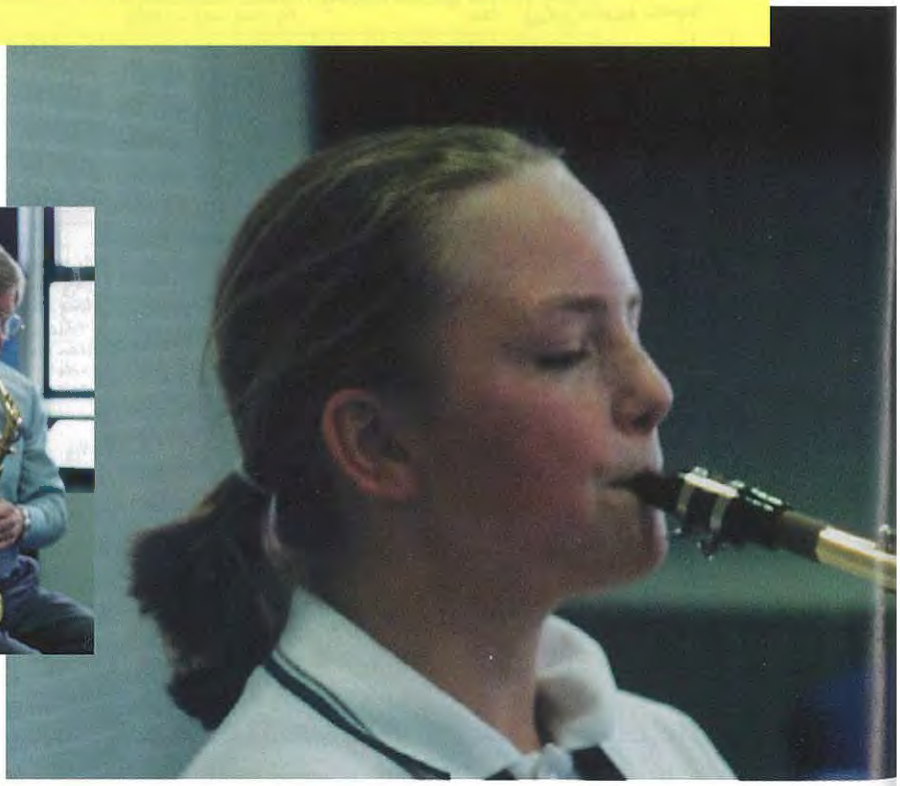
Alice	Brad	David	Dylan
1/9 Beginning to read more difficult (Nate the Great) chapter books competently w/ minimal help... hard work, but can do it!	1/11 Talked about getting off the video game words and stretching yourself me exasperated about his being stuck in the mire of copying w/Harry Took it well - contrite, agreed w/me, made commitment to do more.	1/5 Needing to go to the bathroom a lot. He also noticed this - said he would speak to his parents	1/15 Serious about taking responsibility for his own learning. Stayed in at recess to finish. Only took 5 minutes WELL done
Elizabeth	**Harry**	**Jeri**	**Kathleen**
1/12 off the wall all AM! Popping up and down, interrupting, off-task, noises and other sounds, repeated redirection necessary "respectful behavior" discussed	1/17 Writing story about "Baseball Diamond Mystery" w/Andy - Prominent smudge has it all over it and drawing of what looks like Hitler... nothing said yet, but watching this.	1/8 w/Kleenex reading Annie Sullivan "This is sad" "Do any more people die?" Talked about whether to continue - it's okay to cry/feel sad w/books	1/11 Finished "Kathleen Wore Her..." book Nudged to correct the few misspelled words - 1st time
Lawrence	**Mandy**	**Michael**	**Nancy**
	1/12 Really pleased First day to bring lunch packed at home rather than buying school lunch. "I made it myself!"	1/17 Making a book about fishing trip w/ lift-up flaps showing fish + names of them	
Paul	**Randy**	**Ricky**	**Susan**
1/5 Went to Lawrence's on the bus, but Mom still came by at the end of the day... to check on him?	1/19 On time every day this week... maybe we've licked the tardiness problem?		1/11 Very helpful getting Tory caught up after long illness. Good "teacher", shows how-to, then does own work side-by-side ready to assist "student"
Thomas	**Tory**	**Vince**	**Zoey**
			1/13 Shocked and wanted to share with me what she was reading about M.L. King Jr.

Focusing and recording observations

Post-it notes and focused record sheets

The 'post-it' system could be combined with a record sheet designed to focus observations on particular learning outcomes. Observations made on post-it notes could be transferred to the record sheet which could be used for several students. This example shows how observations of students' performances in relation to four focus outcomes in science might be recorded.[6]

Class	**Activity/project**			
	Focus Outcomes			
Student	Planning Investigations	Conducting Investigations	Using science	Acting responsibly
Beth	15/6 using concept map to plan		17/6 appropriate use of apparatus	17/6 helped student in trouble with bunsen burner
James		17/6 arranged two alternative sequences	17/6 methodol-ogical comparison	



Focusing and recording observations

Class list

Leon Deleuil has developed a checklist for recording observations of students' performances.[7] In this example students make the musical instruments they then play. The aspects of each student's performance (i.e. the assessment criteria) that Leon watches for are shown across the top of the grid. Some of the criteria could be assessed in a structured assessment context; for example, the teacher might set aside time for a student to play the instrument to assess the criterion 'gets a sound'. The criterion 'safe with tools', however, would be assessed as the student makes the instrument. The targeted outcomes are from two levels and two strands of the Australian Science Profile.

Name	Instrument	Gets a sound	Changes note	Makes louder	Safe with tools
Beth	pan pipes	✓✓✓✓✓	✓✓✓✓✓	✓✓✓✓✓	✓✓✓✓✓
Brian	bottle xylophone	✓✓✓✓✓	✓✓✓✓✓	✓✓✓✓✓	
Buckie	guitar	✓✓✓✓✓	✓✓✓✓✓	✓✓✓✓✓	needs more care with knife
Choo	flute	✓✓✓✓✓	✓✓✓✓✓	✓✓✓✓✓	✓✓✓✓✓

estimating and reporting achievement

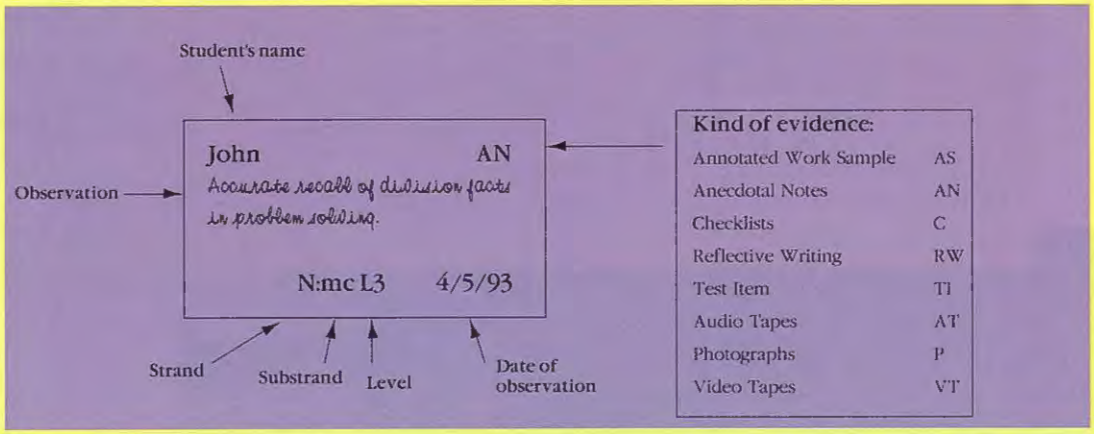
Focused classroom observations provide teachers with evidence that can be used to monitor and report student growth against a progress map. The larger the number of observations, the more evidence a teacher has to estimate a student's level of achievement.

In developmental assessment, growth is monitored against a progress map which describes development within a particular learning area. Focused classroom observations provide teachers with evidence that can be used to monitor and report student growth against such a map. The larger the number of observations, the more evidence a teacher has to estimate a student's level of achievement.

Teachers who use the Bainbridge Island method for recording, reporting and tracking student growth in Writing, make a single on-balance estimate of a student's level of attainment, indicating the estimated level with a date on a tracking sheet (see *ARK Portfolios* p. 21). This system could be adapted for observations in other learning areas. An example of an adaptation for observations of drama performance is shown opposite.

Estimating and reporting locations on a progress map

Lorelle Holcroft and Robyn Coates have developed a system for tracking student growth against the levels of the Queensland Student Performance Standards.[8] Their system is not specific to performance assessment. Each piece of assessment evidence is dated and annotated using two codes. The first indicates the kind of evidence (e.g. AS = annotated work sample). The second is a Student Performance Standard code indicating the strand, substrand and level of work.



These records are later transferred to a Summary Sheet which brings together, in one place, details of the available evidence: the kind of evidence (recording tool), the date of evidence, the strand, substrand and level of performance for *each* observation. Teachers then use this information to make an *on-balance judgement* of a student's level of achievement.

Sub-Strand	Level 1	Level 2	Level 3	Level 4
MENTAL COMPUTATION N: mc	1.15 Use counting and other strategies to mentally solve self-generated or orally presented questions from stories involving small numbers.	2.15 Estimate and calculate mentally, including adding and subtracting numbers to 10 and making extensions based on place value.	3.15 Estimate and calculate mentally, including adding (sum to 100) and subtracting two-digit numbers and multiplying numbers to 10.	4.15 Estimate and calculate mentally, including adding and subtracting most two-digit numbers and multiplying and dividing one digit multiples of ten by one-digit numbers.
Recording tool			AN	
Date evidenced			4/5/93	
WRITTEN COMPUTATION N: wc		2.16 Use a variety of strategies, including regrouping, to assist in adding and subtracting whole numbers which they cannot complete mentally.	3.16 Use understood written methods to add and subtract any whole numbers and amounts of money and to multiply and divide whole numbers by whole numbers to 10.	4.16 Use understood, written methods to add, subtract, multiply and divide whole numbers, money and measures (two places, whole number multipliers and divisors by 10).
Recording tool		TI	AS	
Date evidenced		2/93	3/93	

Estimating and reporting locations on a progress map

The example shown here is an adaptation of the Bainbridge Island method for tracking student growth in writing (see *ARK Portfolios* p. 21). The criteria are taken from Levels 3, 4 and 5 of the Drama Strand of the Arts Curriculum Profile for Australian Schools.

NAME	Level 3 • explores ideas and feelings through drama • explores and uses several drama elements and uses specific skills, techniques and processes such as improvising and role-playing • plans and presents drama for a particular audience or purpose	Level 4 • experiments with ideas and explores feelings to find satisfactory solutions to tasks • selects, combines and manipulates drama elements using a range of skills, techniques and processes • draws upon a range of skills to present drama for a variety of audiences and purposes	Level 5 • uses starting points such as observation, experience and research to express ideas and feelings • structures drama by organising drama elements and applying appropriate skill, techniques and processes • plans, selects and modifies presentations for particular occasions taking into account factors such as purpose, space, materials and equipment
1 Karley A.	9/92	1/93 →	
2 Mackenzie B.	9/92 1/93		
3 Paul B.	9/92 ↔	1/93	
4 Zach C.	9/92	1/93	
5 Lucas C.	9/92 1/93		
6 Win D.		9/92	1/93
7 William D.	9/92 1/93		
8 Tiffany E.			11/92
9 Clayton H.	11/92 →		
10 Nicole H.		12/92	1/93
11 Julie H.	9/92 2/93		
12 Molly J.	9/92 1/93		
13 Kirstin L.		9/92	2/93
14 Michelle M.		9/92	
15 J.J. M.	9/92 2/93 No growth!		
16 Robbie O.	9/92 2/93		
17 Adrian S.	9/92		2/93
18 Ryan S.	9/92	2/93 →	
19 Patrick S.	9/92	2/93	
20 Alex W.	9/92	2/93 w/help	
21			
22			
23			
24			
25			

1 Terry Spottiswood, Waldau Primary School, Victoria, Australia.

2 Killen, R. (1993) *Assessment and Evaluation*: Newcastle: University of Newcastle, Faculty of Education, p. 94.

3 Beyer, A (1993) 'Assessing Students' Performance Using Observations, Reflections, and Other Methods' in Webb, N. & Coxford, A. *Assessment in the Mathematics Classroom* Virginia: National Council of Teachers of Mathematics, Inc., p. 119.

4 Olsen, K., Adams, G., Grace, N. & Anderson, P. (1994) *Using the Mathematics Profile*, Carlton: Curriculum Corporation, p. 84.

5 Campbell Hill, B. & Ruptic C. (1994) *Practical Aspects of Authentic Assessment* Norwood MA: Christopher-Gordon Publishers, Inc., pp. 55-60.

6 Adapted from Deleuil, L. and Malcolm, C. (1994) *Using the Science Profile* Carlton: Curriculum Corporation, p. 73.

7 Deleuil, L. and Malcolm, C. (1994) *Using the Science Profile* Carlton: Curriculum Corporation, p. 47.

8 Holcroft, L. and Coates, R. *Monitoring Learning in Mathematics* Brisbane: Mount Coot-tha School Support Centre.

classroom performance assessment 'events'

Teachers sometimes arrange for students to make presentations or to give performances with the intention of assessing those presentations and performances. Examples include: a performance of an original music composition, a role play conversation in Japanese, a dance or drama production, an assessment of running technique, and an oral presentation of the results of a local history investigation.

Performance assessments of this kind are different from informal classroom observations made in the course of teaching (see pages 9-18). They also differ from structured assessments designed for high-stakes contexts, where special efforts are made to ensure the comparability of assessments from student-to-student, assessor-to-assessor and school-to-school (see pages 37-45).

Teachers use performance assessment events, rather than informal observations, to collect information they would be unlikely to see in passing, and to collect it efficiently.

These are *planned* assessment occasions for the observation of students *engaged in an activity*. The teacher selects a context for the assessment and decides, in advance, on a method for judging students' performances. These judgements provide evidence about 'doing' outcomes in a learning area.

Assessment criteria are developed to focus teachers' observations on particular kinds of evidence, and a system for recording observations is used.

planning performance assessment events

The usefulness of a presentation or performance as a source of evidence about 'doing' outcomes in a learning area depends on:

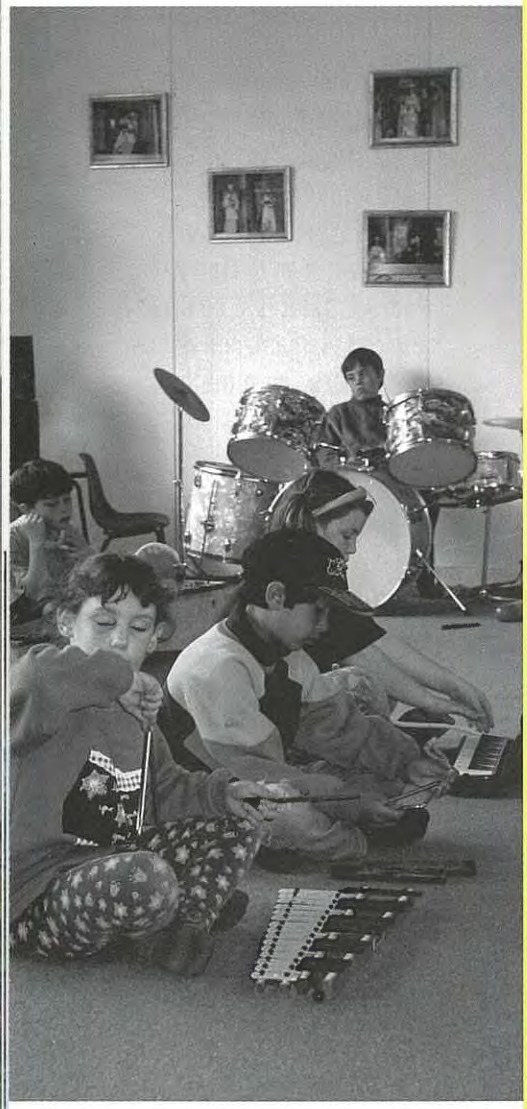
- how well the evidence addresses the instructional goals, or 'outcomes', of the learning area (relevance);
- how well the performance task samples the range of learning area outcomes (coverage);
- how 'fair' the task is to students from different language, cultural and socio-economic backgrounds; and
- how well the task allows students to show their strengths and weaknesses.

The manageability of performance tasks in the classroom also is an important consideration. A task which provides excellent information about a student's level of achievement but requires a teacher's attention to be directed exclusively to one student for an extended period of time may be impractical in a classroom context.

Classroom performance assessment 'events' are *planned* assessment occasions for the observation of students *engaged in an activity*. The teacher selects a context for the assessment and decides, in advance, on a method for judging students' performances. These judgements provide evidence about 'doing' outcomes in a learning area.

addressing outcomes

When assessing student performances it is important to be clear about the outcomes to be addressed. Tasks can then be structured appropriately and observations focused for the collection of evidence including outcomes from different levels of a progress map. An assessment event is most useful when it allows evidence to be collected about a range of outcomes.



Planning performance assessment events

Addressing outcomes

On completing a unit of work, Terry Spottiswood's students perform a speaking task as their final assessment.[1] They introduce books they have written and report the process to an audience of peers, younger students and parents. Terry focuses her assessment on outcomes taken from the 'linguistic structures and features' sub-strand of the 'speaking and listening' strand of the English profile. Can students control *'most linguistic structures and features of spoken language for interpreting meaning and developing and presenting ideas and information in familiar situations'*? Terry looks particularly for students' abilities to:

- arouse and maintain audience interest;
- use complex sentences in spoken language;
- vary tone, pitch and pace; and
- overcome nervousness.

She makes an on-the-spot evaluation using a system of ticks to indicate degree of competence: one tick - good, two ticks - very good, three ticks - excellent, four ticks - outstanding. Terry also leaves space on her record sheet for other comments.

Name	Arouses and maintains audience interest	Uses complex sentences	Varies tone, pitch and pace	Has second try to make appropriate	Other comments
Jody	✓✓	✓✓	✓✓	–	
Nam Huong	✓✓	–	✓✓	✓	greatly improved really nervous
Tom	✓✓✓✓	✓✓	✓	–	stunning
Gino	✓	✓	✓	–	well sequenced. good points

ensuring a depth of evidence

Teachers use presentations and performances as opportunities to collect information about students' underlying skills and abilities. The particular topics on which students make presentations, the particular pieces of music they play, or the particular dance routines they perform usually are less important than the underlying skills and abilities about which those performances provide evidence. The more opportunities a teacher has to observe a student's performance, the more reliable will be their judgement of the student's underlying ability. Observations of student performance need to occur on a *number of occasions*.

Observations of performances also need to occur in a *range of contexts*. A teacher assessing oral language might observe a student in different kinds of oral performance - small group discussions, formal speeches, with a range of audiences - before she is prepared to make a judgement about the student's level of speaking attainment.

In thinking about this range of contexts, teachers need to consider whether assessments are of the *same kind* of underlying skill or ability. Does it make sense to infer something general about a student's 'ball-handling' skills from performances in contexts as different as basketball, tennis, and soccer, for example?

considering fairness

It is also important to be aware of the ways in which gender, cultural background and personality can influence a student's performance. Teachers need to consider the contexts of performance events and the assessment criteria to make sure they are 'fair' indicators of achievement.

Observations of student performance need to occur on a *number of occasions* and in a *range of contexts*.

Planning performance assessment events

Considering fairness

Students' performances can be influenced by characteristics of the assessment situation irrelevant to the outcomes being assessed.

Some examples:

gender

Do girls perform as well in physical activities if they are playing in mixed teams?

Steve Quartermaine found that girls generally display lower levels of physical skill when they are assessed in mixed teams.[2]

cultural background

Do some cultural practices disadvantage students in performance assessments?

An oral performance assessment may have 'keeping eye contact' as a criterion for judgement. This may be inappropriate for cultural groups for whom making eye contact is not encouraged.

allowing for a range of levels of performance

Performance assessment events need to be open-ended enough to allow low-achieving students to experience success, while challenging other students and allowing them to demonstrate higher levels of achievement. An assessment event is most useful when it allows evidence to be collected across several outcome levels.

In the example below, Year 6 students' skill at estimating and measuring capacity and volume is assessed against five levels of performance.

Planning performance assessment events

Allowing students to show a range of levels of performance

The Toronto Board of Education Benchmark task, 'Pouring Water', is sufficiently open-ended to allow the task to be attempted in a variety of ways reflecting different levels of performance.[3] The Board defines five levels of performance on the task and shows the percentage of students performing at each level.

RANGE OF RESPONSES	HOLISTIC SCORING CRITERIA
6%	**LEVEL FIVE** The student uses correct methods to measure accurately and efficiently with water. Measures are made with the containers on the level and part measures are monitored and recorded. Measures are done quickly and with confidence. The student is fluent in metric units and includes appropriate units without prompting. The student gives close estimates.
42%	**LEVEL FOUR** The student may experiment to find correct methods to measure with water usually arriving at correct results. The student may be inconsistent in methods used to measure and units chosen. There may be an initial confusion concerning the task but the student may learn as the task proceeds. Initial estimates are usually in appropriate units but are not close.
29%	**LEVEL THREE** The student has a weak understanding of how to measure with water. Measures are made perhaps without awareness of the need for accuracy and completeness; methods may be changed as the student proceeds with the task. Units are confused at times or not used appropriately. There may be some inefficiency in the measuring such as losing track of "part" measures. Estimates indicate an incomplete understanding of underlying concepts. Some coaching is required.
11%	**LEVEL TWO** The student uses inefficient methods to measure or changes methods to no apparent advantage. Measures are inaccurate and incomplete. The student is confused about units and may count the units on the graduated cylinder. Estimates indicate very little understanding of underlying concepts. The student needs to be coached.
12%	**LEVEL ONE** The student does not understand the problem or forgets the assigned task.

judging and recording

The criteria for judging student performances and the way in which judgements are recorded depend on the learning area, the outcomes addressed and the immediate assessment context.

The examples on pages 23-31 show assessment criteria and recording methods that have been used in different kinds of performance assessment events. In these examples judgements are made either analytically or holistically by teachers, peers, or self.

Judging and recording

Analytic rating (Self-assessment)

Lorelle Holcroft and Robyn Coates suggest that students evaluate their completion and understanding of set tasks and participation in a group problem solving activity.[4] These self-assessments can form the basis for subsequent teacher conferencing.

Name:

Activity:

I worked in a group of ..... people

The other members of my group were .....

.....

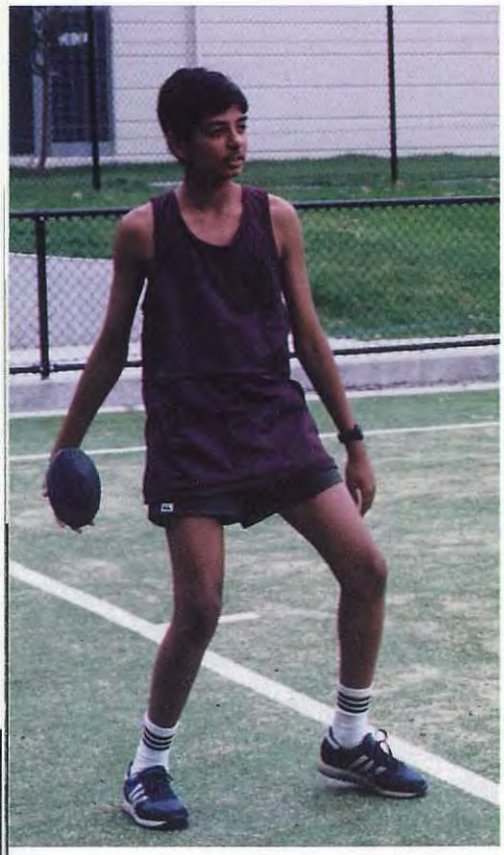
While solving this problem, I

• felt confident	YES		NO
• enjoyed what I was doing	NOT AT ALL	SOMETIMES	VERY MUCH
• made suggestions to try	NOT AT ALL	SOMETIMES	VERY MUCH
• listened to what others in my group had to say	NEVER	SOMETIMES	WHEN APPROPRIATE
• participated in the group discussions	NEVER	SOMETIMES	ALWAYS
• helped others in my group	NEVER		WHEN I COULD
• gave up easily/kept on trying	GAVE UP EASILY		KEPT ON TRYING

When I was working through this problem, I was able to show my teacher that I knew.....

Apart from the other members of my group, I talked about this problem to

On a five point scale [A B C D E], for my report I think I should receive



Judging and recording

Analytic rating (Self-assessment)

The Connecticut Common Core of Learning Assessment project has developed a self-evaluation form for assessing participation in a group activity.[5] Students circulate their self-ratings to each person in the group for his or her review and signature at the completion of the activity. If any member or the group disagrees with the student's self-ratings, the reasons for the disagreement are discussed. The student then decides whether to change the original rating.

Student name _____

Student I.D. number _____

	Fill in one			
A. GROUP PARTICIPATION	Almost always	Often	Sometimes	Rarely
1. Participated in group discussion without prompting	O	O	O	O
2. Did his or her fair share of the work	O	O	O	O
3. Tried to dominate the group, interrupted others, spoke too much	O	O	O	O
4. Participated in the group's activities	O	O	O	O
B. STAYING ON THE TOPIC	Almost always	Often	Sometimes	Rarely
5. Paid attention, listened to what was being said and done	O	O	O	O
6. Made comments aimed at getting the group back to the topic	O	O	O	O
7. Got off the topic or changed the subject	O	O	O	O
8. Stayed on the topic	O	O	O	O
C. OFFERING USEFUL IDEAS	Almost always	Often	Sometimes	Rarely
9. Gave ideas and suggestions that helped the group	O	O	O	O
10. Offered helpful criticism and comments	O	O	O	O
11. Influenced the group's decisions and plans	O	O	O	O
12. Offered useful ideas	O	O	O	O
D. CONSIDERATION	Almost always	Often	Sometimes	Rarely
13. Made positive, enconraging remarks about group members and their ideas	O	O	O	O
14. Gave recognition and credit to others for their ideas	O	O	O	O
15. Made inconsiderate or hostile comments about a group member	O	O	O	O
16. Was considerate of others	O	O	O	O
E. INVOLVING OTHERS	Almost always	Often	Sometimes	Rarely
17. Got others involved by asking questions, requesting input, or challenging others	O	O	O	O
18. Tried to get the group working together to reach group agreements	O	O	O	O
19. Seriously considered the ideas of others	O	O	O	O
20. Involved others	O	O	O	O
F. COMMUNICATING	Almost always	Often	Sometimes	Rarely
21. Spoke clearly, was easy to hear and understand	O	O	O	O
22. Expressed ideas clearly and effectively	O	O	O	O
23. Communicated elearly	O	O	O	O
G. OVERALL EXPERIENCE	Almost always	Often	Sometimes	Rarely
24. This group helped me to iinprove my understanding of the problems and the ways of solving them more than if I had worked alone	O	O	O	O
25. Working with the group was an enjoyable experience	O	O	O	O

Judging and recording

Analytic rating (Self-assessment)

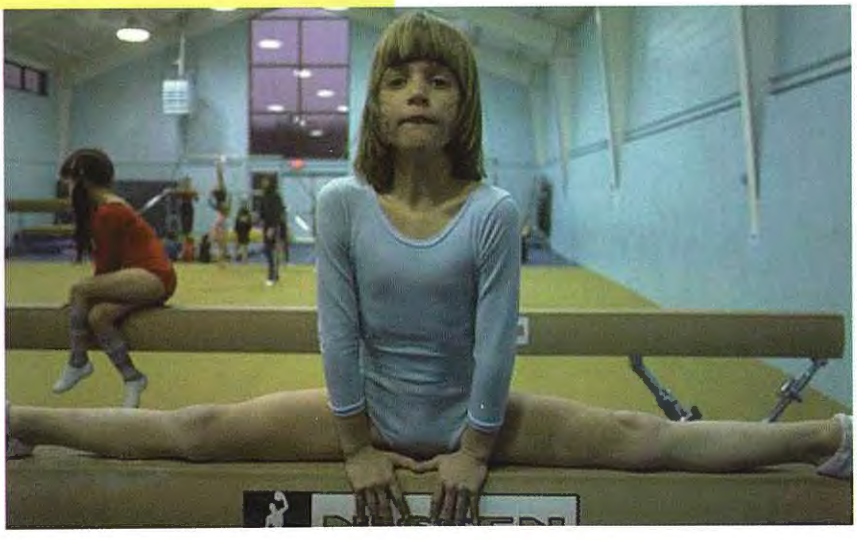
Ian Fox provides this example of a general self-evaluation form for young school gymnasts.[6] Students rate their own performance on a five point scale 'Low' to 'High' on seven pieces of apparatus. They are invited to comment on their enjoyment of gymnastics and to reflect on where they need further practice. Students could be asked, on the basis of these assessments on each piece of apparatus, to make an on-balance judgement of their overall gymnastics performance.

Children's Gymnastics

	LOW				HIGH
Beam					
Trampoline					
Vault					
Ropes					
Tumbling Mat					
Rings					
Crash Mat					

Signed: _____

Teacher: _____



Judging and recording

Analytic rating (Teacher and Self-assessment)

Ian Fox provides this example of a combined teacher and self-evaluation form for music performance using tuned percussion instruments.[7] Teachers and students rate performances on five criteria using a 3-point scale: 'Still Trying', 'Has Mastered', 'Competent'.

Music Playing Skills

	Still Trying		Has Mastered		Competent	
	student	teacher	student	teacher	student	teacher
TONE		✓	✓			
TIMING					✓	✓
USING THE 'BEATERS'					✓	✓
READING MUSIC					✓	✓
SMOOTHNESS				✓	✓	

Judging and recording

Analytic rating (Teacher assessment)

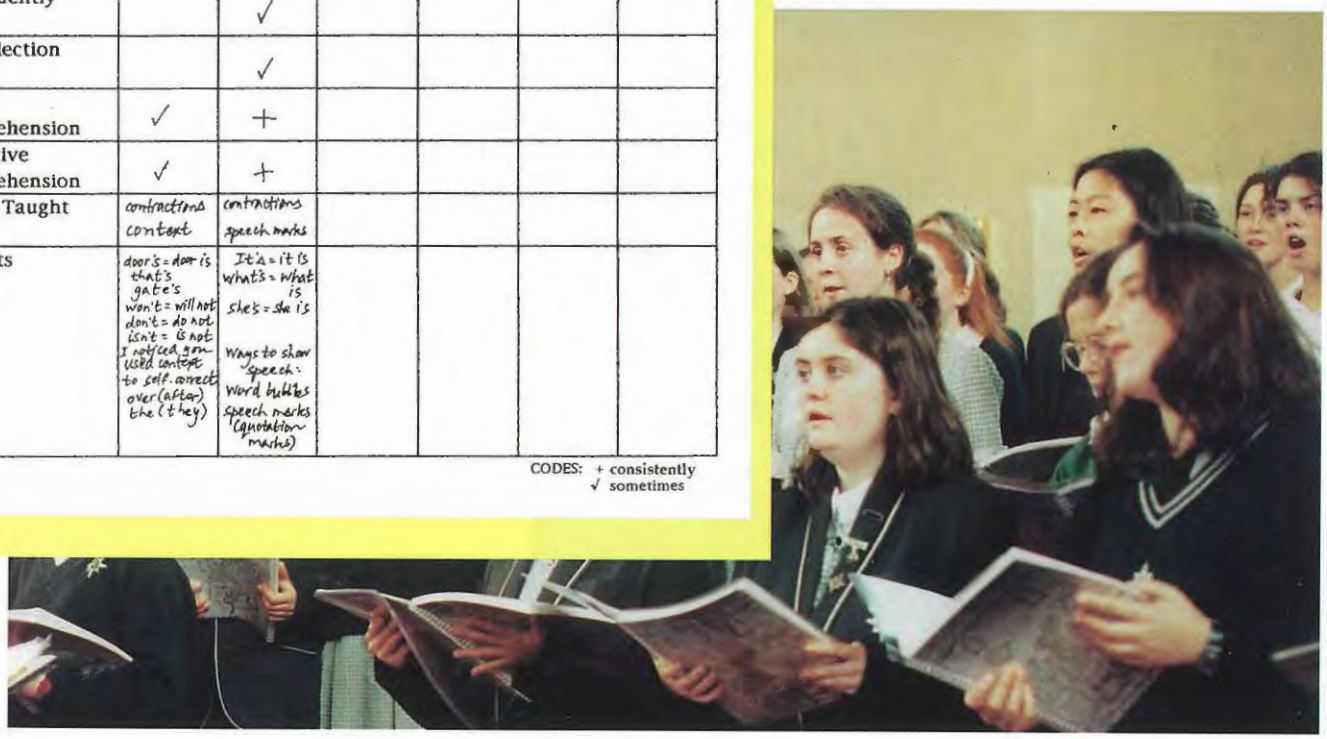
Structured conferences and interviews are another kind of assessment event which can be used to assess specific learning outcomes. A structured reading conference, for example, gives teachers an opportunity to listen to a student read aloud from a selected text and to assess that student's reading skills. Cindy Fulton has developed the following form for rating students' reading strategies, fluency and inflection, and reading comprehension.[8] A tick indicates that the student displays the skill 'Sometimes', a plus indicates that the skill is displayed 'Consistently' during the conference. When teachers use this rating form in different reading conferences with a student, they are able to document and see improvement over time.

Reading Conference Record

Student_____

	Date: 12/8	Date: 12/17	Date:	Date:	Date:	Date:
	Title: Two Little Dogs (G)	Title: Sleeping out (F)	Title:	Title:	Title:	Title:
Level Appropriate	Instructional Frustration ✓	Instructional Independent				
Strategies: Whole idea						
Picture clues	✓					
Pattern						
Sight words	✓	✓				
First letter						
Decodes	✓	✓				
Context clues	✓	✓				
Skip/return						
Rereads	✓	✓				
Reads Fluently		✓				
With Inflection		✓				
Literal Comprehension	✓	+				
Interpretive Comprehension	✓	+				
Strategy Taught	contractions context	contractions speech marks				
Comments	door's = door is that's gate's won't = will not don't = do not isn't = is not I noticed you used context to self-correct over (after) the (they)	It's = it is what's = what is she's = she is Ways to show speech: Word bubbles speech marks (quotation marks)				

CODES: + consistently
√ sometimes



Judging and recording

Analytic rating (Teacher assessment)

In this analytic rating scheme for a group music performance, eight aspects of each participating student's performance are assessed. Each aspect (criterion) is judged on a 4-point scale: 'not evident', 'limited competence', 'competent', 'very competent'. Space is provided for the teacher to make comments.

ASPECTS OF MUSIC PERFORMANCE	Not Evident	Limited Competence	Competent	Very Competent	Comments
1 Performance technique					
2 Performance control					
3 Quality of tone					
4 Variation of tone					
5 Articulation and phrasing					
6 Skill in performing as a member of the group					
7 Individual 'musicality'					
8 Presentation					

Judging and recording

Holistic rating (Teacher assessment)

This guide for assessing students' performances on a DART small group discussion task provides teachers with described levels of performance that match Levels 2, 3, 4 and 5 of the English Profile for Australian Schools.[9] Teachers observe individual students working in a small group and make an on-balance judgement of their level of performance.

Level 5

- implicitly or explicitly directs and summarises group discussion and co-ordinates decisions about organising performance
- views are relevant and persuasive, and reflect some awareness of how different people's views are influenced
- may pick up the implications of other group members' contributions and articulate them
- co-ordinates decisions and organises tasks where appropriate

Level 4

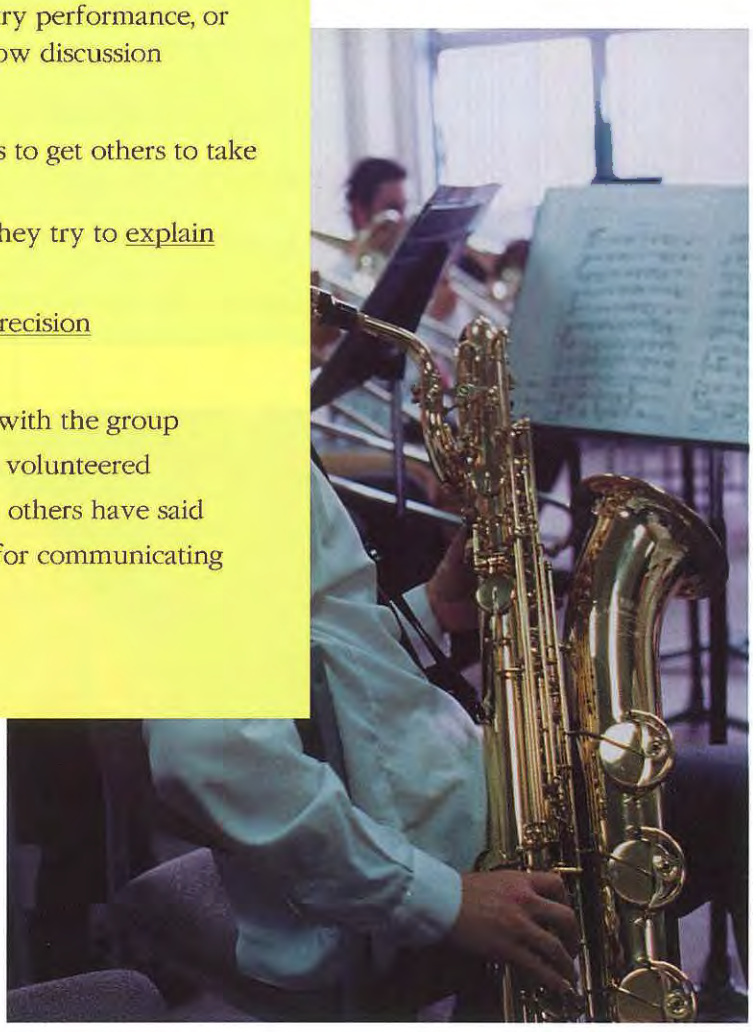
- takes a lead in organising the group discussion, encourages others and directs and summarises progress
- offers own opinions forcefully (but does not necessarily dominate)
- is articulate and persuasive in the way they express themselves
- plays a substantial part in organising the poetry performance, or scribing and reporting the TV roles or pet show discussion

Level 3

- initiates some parts of the discussion and tries to get others to take part
- offers own opinions about the topic, which they try to explain
- listens responsively to others' points of view
- language has some flavour of originality or precision

Level 2

- understands what the task is and cooperates with the group
- views about the topic are elicited rather than volunteered
- asks questions to help in understanding what others have said
- uses language which is simple and adequate for communicating ideas



Judging and recording

Analytic rating (Teacher assessment)

Neryl Jeanneret provides the following example of an analytic rating scale for drama performance.[10] Eight aspects of performance are assessed using a 4-point scale from 'strongly disagree that the statement is descriptive' to 'strongly agree that the statement is descriptive'.

Rating scale
- SD Strongly disagree that the statement is descriptive.
- D Disagree that the statement is descriptive.
- A Agree that the statement is descriptive.
- SA Strongly agree that the statement is descriptive.

ASPECTS OF DRAMA PERFORMANCE	SD	D	A	SA
1 The performance is well-rehearsed.				
2 The performance is not mechanical like a robot.				
3 The item is within the technical capabilities of the performer.				
4 The performance is confident.				
5 The style/genre is successfully portrayed.				
6 The performer is in control of the performance.				
7 The performance is 'artistically' successful.				
8 There is original input into the interpretation.				

Judging and recording

Analytic rating (Teacher assessment)

The Student Needs Assessment Procedures (SNAP) have been developed to assist teachers to make formal assessments of the oral and written work of students from non-English speaking backgrounds.[11] Oral language assessment activities for secondary school students include: group discussion, formal debate, interview, meeting procedure, oral argument, oral description and oral narrative. In the assessment of performances on a SNAP oral argument task, teachers make judgements on a range of criteria. Explanatory notes are provided to assist assessment.

Secondary Oral Language Assessment Activity -ORAL ARGUMENT

Name of student: Year level: Class: Date:
Task: Context:

Criteria (Tick appropriate box)	Very competent	Competent	Limited competence	Not evident	Comments
Schematic Structure *Did the student:*					
• make an opening statement previewing the issue					
• make a position statement ie Thesis					
• present relevant arguments to support the position statement					
• support the arguments with appropriate evidence					
• anticipate and refute an opposing viewpoint (optional)					
• summarise evidence & make appeal for action (if appropriate)					
Language Features					
Participants *Did the student:*					
• provide detailed information about participants					
• use topic specific vocabulary					
Processes and Circumstances *Did the student:*					
• use a range of verbs (processes)					
• use appropriate tense					
• use a range of circumstances					
Tenor (Communicative Effect) *Did the student:*					
• vary use of person					
• use language in a personal and interactive way if making an appeal for action or use language neutral/impersonal in tenor if arguing 'analytically'					
• use modality to express obligation (in an appeal for action) or certainty (in an appeal to logic)					
• speak clearly and audibly					
• explain unfamiliar terms					
• respond to audience					
• speak fluently					
• use body language to establish/maintain rapport					
Cohesion and complexity of language *Did the student:*					
• use reference items					
• use a range of conjunctions					
• use complex sentences					
Accuracy *Did the student:*					
• use grammar accurately and without omissions					
- verbs (processes)					
- articles					
- prepositions					
- word order/sentence structure					
• use agreement accurately					
• approximate English pronunciation, stress and intonation patterns					
• self-correct					
Strategies, skills & support *Did the student:*					
• use information gathering/research skills					
• complete the task independently (ie with minimal support)					
• refer to pictures, realia to enhance meaning (optional)					

General comments (including attitude)

Global Rating: (Circle) 1 2 3 4 5 highest

estimating and reporting locations on a progress map

Performance assessment events provide evidence that can be used to estimate students' locations on a progress map.

Performance assessment events provide evidence that can be used to estimate students' locations on a progress map. This evidence can take a variety of forms, including teachers' records of behaviours demonstrated during an assessment event, analytic ratings of various aspects of a performance, holistic ratings of entire performances, and records of scored student responses.

When teachers keep records of behaviours/outcomes listed at different levels on a progress map, they can use these records to make on-balance decisions about students' overall levels of achievement (see below).

Decisions about students' levels of achievement also can be made on the basis of analytic and holistic ratings of performances, particularly if the rating levels are aligned directly with described levels on a map. In the example on page 33, students' soccer performances are rated on two criteria (i.e. analytically) using a set of three described levels of performance constructed to correspond to Levels 2, 3 and 4 of a Western Australian Physical Education progress map.

Estimating and reporting

Records of observed behaviours

Cindy Ruptic has designed a form for her reading conferences that includes the Bainbridge Island Reading Levels (Pre-conventional, Emergent, Developing, Beginning, and Expanding).[12] During the conference Cindy makes detailed notes on the student's reading skills. She also circles performance indicators. She then makes an on-balance judgement of the student's performance. In this example Cindy has circled all indicators at the 'Developing' stage, and some at the 'Beginning' stage. Cindy's overall assessment is that this student is functioning at the 'Developing' level.

Date: 12/2 Title(s): Two Little Dogs (Wright-G) 16 pages very difficult!

EVALUATION OF READING	LEVEL:	Frustration ← Instructional → Independent
PRE-CONVENTIONAL	Print concepts; Letter names; Focuses on pictures; Responds; Chooses/has favorites	
EMERGENT	Pictures to tell story; Names/words in context; Letter sounds; Memorizes pattern; Rhyme	
DEVELOPING	Self=reader; Word concepts; Print+pictures; Simple words; Pattern books; Retells main idea	
BEGINNING	Level appro.; Sight words; Self-corrects; Punctuation; Phonics; Sentence structure; Context	
EXPANDING	Many strategies; Short chapter books; Silently; Expression; Retells; Connects to experience	

Strategies, Questions & Evidence	Pictures	Pattern	Phonics	Sight Words	Context	Miscues	SC	TG	I
Halting; working through text laboriously, but w/ good comprehension door's = door is that's = that is gate's = gate is won't = will not don't = do not isn't = is not	gate's mouse rabbit		look door's that's good little bad him he way open	the we ran run	two	over (after) the (they) beginning to use the context to self-correct	v v		

*Strategy/Assignment: 's and n't contractions / using context clues "I noticed" you're using text to help... con: with text = text

© 1993 C. Ruptic, Bainbridge Island, WA CODES: SC=Self-Correct, TG=Teacher-Given, I=Ignored

Estimating and reporting

Analytic rating

Teachers using this Monitoring Standards in Education marking guide rate two aspects of each student's performance.[13] The described levels of performance correspond to levels 2, 3, and 4 of the Western Australian Outcome Statements. Thus these two ratings could be used directly to estimate a student's level of achievement in this area of the physical education curriculum.

Soccer skill	Outcome level 2	3	4
Sending Kicking	**Inconsistent use of correct technique** 1. Kicks with toe 2. Poor transfer of weight 3. Looks awkward 4. Inaccurate	**Correct technique and control without pressure** 1. Kicks with inside of foot 2. Limited transfer of weight 3. Maintains balance 4. Often inaccurate	**Correct technique and control under pressure** 1. Uses arms for balance 2. Efficient transfer of weight 3. Controls distance and direction 4. Accurate
	Outcome level 2	3	4
Receiving Trapping	**Controls the ball rarely** 1. Foot rigid 2. Controls ball without pressure sometimes	**Controls the ball with correct technique without pressure** 1. Gives with the foot 2. Maintains balance 3. Controls ball on the ground	**Controls the ball with correct technique under pressure** 1. Traps on the run 2. Controls most passes 3. Controls ball in flight and on the ground

Students' levels of attainment on a progress map also can be estimated from scored records of their performances on assigned tasks. In the performance assessment event described below, teachers assign scores to children's performances on seven separate fine motor tasks, each scored 0 to 3.

Children's total scores (0 to 21) provide estimates of their locations on a described map of developing Fine Motor skills (see page 35).

Estimating and reporting ***Scored responses***

Carol Mardell and Dorothea Goldenberg have developed a screening instrument known as the Developmental Indicators for the Assessment of Learning (DIAL) to identify pre-kindergarten children in need of follow-up services and more detailed diagnostic evaluation.[14]

As part of DIAL, children are administered a set of tasks designed to assess their Fine Motor skills (matching shapes, building with blocks, cutting card, copying shapes, copying letters, touching thumb to fingers, and clapping hands). Assessors observe each child and assign a score of 0 to 3 to the child's performance on each task (see scoring for the Touching Fingers task below).

Each child receives a total score of between 0 and 21 on the set of Fine Motor tasks. This score provides an estimate of the child's position on a map of developing Fine Motor skills defined by these seven tasks (see page 35). A child with a score of 11, for example, would typically have attained a level of Fine Motor skill which would allow him to match at least some shapes, build simple three-block towers, cut along straight lines, and clap hands twice.

Touching Fingers

a) Operator says: **OK (child's name), watch what I do.**
b) Operator touches each finger on one hand to the thumb on the same hand at the rate of one per second.
c) Operator then says: **Now you do it.**
d) Operator says **Now try your other hand** regardless of whether child succeeded with first hand.

Scoring

3 child touches thumb to fingers in sequence (going from either direction) with both hands;
2 child touches thumb to fingers in sequence (going from either direction) with one hand;
1 child touches thumb to all fingers regardless of sequence with both hands.

Developing fine motor skills

19

18

Claps hands approx. ten times gradually increasing in tempo.

17

Touches thumb to fingers in sequence on each hand.

16

15

Builds three towers:

14

Touches thumb to fingers in sequence on one hand only.

13

Cuts along curved line with scissors.

12

Touches thumb to all fingers but not in sequence.

11

Cuts along straight line with scissors.

10

9

Claps hands twice in front of body, waist height.

8

7

Builds two 3-block towers:

6

5

Cuts card with scissors, but not along lines provided.

4

Builds one 3-block tower:

3

Places at least three of the following shapes on top of pictures of these shapes:

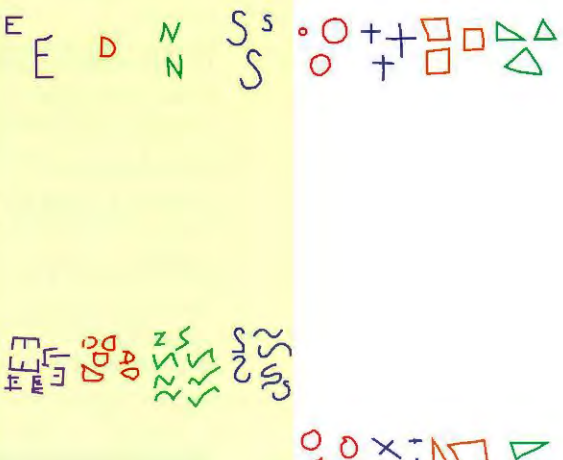
2

1

Attempts to copy the letters

E D N S

Attempts to copy the shapes



1 Terry Spottiswood, Waldau Primary School, Victoria.

2 Conversation with Steve Quartermaine, Monitoring Standards in Education, Department of Education, Western Australia, 1995.

3 Toronto Board of Education, Benchmarks 'Pouring Water' Task.

4 Holcroft, L. and Coates, R. *Monitoring Learning in Mathematics*, Brisbane: Mount Coot-tha School Support Centre.

5 Judith Collison (1992) 'Using Performance Assessment', *The Arithmetic Teacher*, Vol. 39, No. 6, p. 45.

6 Ian Fox, Bucklands Beach Intermediate School, Auckland, New Zealand.

7 as above

8 Campbell Hill, B. and Ruptic, C. (1994) *Practical Aspects of Authentic Assessment*, Norwood, MA: Christopher-Gordon Publishers, Inc. p. 109.

9 Forster, M. Mendelovits, J. & Masters, G.N. (1994) *Developmental Assessment Resource for Teachers DART English*, Camberwell: Australian Council for Educational Research, p. 84.

10 Killen, R. (1993) *Assessment and Evaluation*, Newcastle: University of Newcastle, Faculty of Education, p. 95.

11 Mincham, L. (1992) *ESL Student Needs Assessment Procedures R-12*, Interim Publication, South Australian Education Department.

12 Campbell Hill, B. and Ruptic, C. (1994) *Practical Aspects of Authentic Assessment*, Norwood, MA: Christopher-Gordon Publishers, Inc. p. 113.

13 Monitoring Standards in Education, Department of Education, Western Australia.

14 Mardell, C. and Goldenberg, D.S. (1972) DIAL Developmental Indicators for the Assessment of Learning, DIAL Inc. Illinois.

when comparability matters

It is sometimes important that assessments of students' performances are comparable from student to student, from assessor to assessor, and from school to school. In high-stakes settings students' performances can influence admission to courses, scholarship offers, or the award of certificates. Examples include final year oral examinations in languages, instrumental performances for selection into music programs, dance scholarship examinations, and music certificate examinations. In these situations high levels of student-to-student and assessor-to-assessor comparability, or *fairness*, are sought.

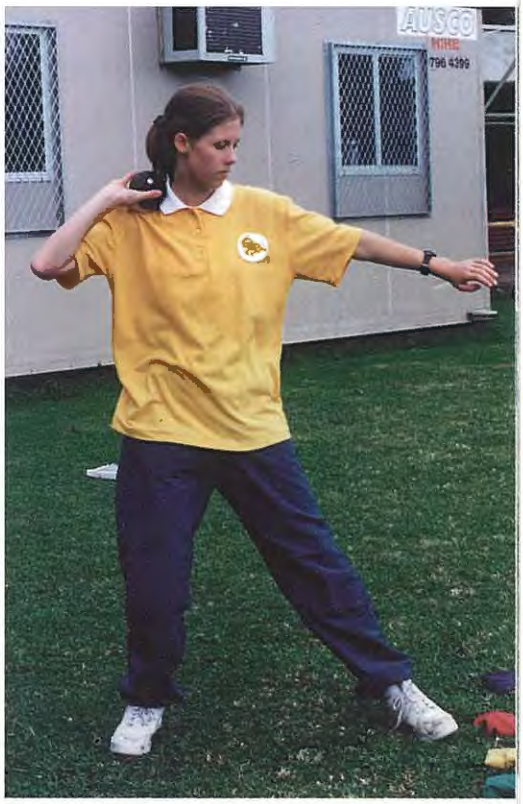
Special efforts also are made to ensure the comparability of performance assessments in programs which monitor levels of student achievement across education systems and over time. Examples include the use of science performance tasks in national assessment programs such as the U.S. National Assessment of Educational Progress (NAEP) and the U.K. Assessment of Performance Unit (APU).

In situations in which comparability is important, special consideration must be given to:

planning performance assessments

- curriculum match - the extent to which performance assessments address a broad range of learning goals;
- fairness - the extent to which performance assessments are fair to all groups of students;
- standardisation - the extent to which the conditions of assessment are similiar from student to student;
- generalisability - the extent to which assessments can be generalised beyond the particular tasks on which they are based;

In high-stakes settings where students' performances can influence admission to courses, scholarship offers, or the award of certificates, high levels of student-to-student and assessor-to-assessor comparability, or *fairness*, are sought.



Examples of contexts in which performance assessment requires student-to-student, assessor-to-assessor and school-to-school comparability:

End of school examinations in

- Dance styles
- Drama solo performance
- Music group performance
- Languages Other Than English (LOTE) orals

Music Examinations

Diving Championships

Music auditions for admission to Arts College

State or National assessments involving performances

(e.g. WA Monitoring Standards in Education - Speaking assessment, U.S. National Assessment of Educational Progress, and U.K. National Curriculum Assessment)

When comparability matters, special consideration must be given to:

- curriculum match;
- fairness;
- generalisability;
- standardisation;
- inter-rater reliability; and
- objectivity.

judging performances

- inter-rater reliability - the extent to which different assessors are consistent in their use of assessment criteria; and

summarising and reporting achievement

- objectivity - the extent to which procedures for making on-balance estimates of students' levels of attainment are specified.

planning performance assessments

When comparability matters, performance assessment provides information not only about what students know and can do, but also for making decisions about what grades they receive, and whether they are to be awarded scholarships or gain entry to university courses, for example. If fair decisions are to be made, then performance assessments must be valid and reliable.

curriculum match

Because high-stakes performance assessment can narrow and possibly distort student learning there must be a good match between assessment tasks and the broad range of curriculum and instructional goals. One method for ensuring this match is to have a panel of experts review assessment tasks.

fairness

It is important that performance assessment tasks do not unfairly advantage or disadvantage individual students or particular groups of students.

Students must have the necessary physical resources to complete performance tasks. A task that requires running water or electricity outlets, for example, might be excluded from a science performance assessment because it could disadvantage students in some rural environments.

Students also must be assessed in contexts which are not so unfamiliar as to adversly affect their performance. A student who is a competent public speaker may be intimidated by having to use a microphone for the first time during a speaking assessment, for example.

During performance assessments students are vulnerable to immediate audience feedback in a way that they are not in other kinds of assessment (e.g. paper and pen, portfolio or project assessment). For this reason, performance assessments, in particular, should occur in supportive environments. Performances may be affected by who is doing the assessment (e.g. students from particular cultural backgrounds may feel uncomfortable being assessed by adults from

The greater the requirement for comparability, the more carefully:

- assessment tasks are structured;
- assessment criteria are specified; and
- the process for estimating students' levels of attainment is specified.

Planning performance assessment

Curriculum match

In the Western Australian Monitoring Standards in Education (MSE) program a team of physical education teachers and measurement specialists reviews the proposed performance assessments to ensure a match between the tasks and the range of curriculum and instructional goals. The assessments are then thoroughly pilot tested on groups of students and teachers.

Planning performance assessment **Ensuring fairness**

In international studies it is important that assessment tasks do not unfairly advantage or disadvantage students from different cultural, language or socio-economic groups.

Performance tasks designed for the IEA Third International Mathematics and Science Study (TIMSS) were screened by an international panel to ensure the tasks were equally meaningful to students from different backgrounds. Tasks also were thoroughly pilot tested before being accepted.

Some tasks were excluded because they required access to particular resources, like hot and cold running water and electricity outlets, which may not have been available in some schools. Others were excluded if they were seen to disadvantage urban students (e.g. tasks requiring agricultural knowledge).

another cultural background), by who is present during the assessment (e.g. students may perform poorly before unsympathetic peers), and by who is included in the performance task (e.g. girls may perform less well in mixed than in single sex contexts).

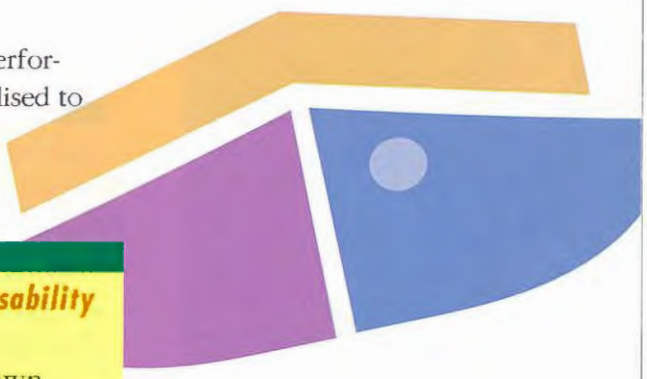
Expert review of proposed assessment tasks and careful trial testing of tasks can help to ensure fairness and lack of bias.

generalisability

Although assessments of student achievement are based on observations of performances on specific tasks, the purpose and interest usually is in some more general underlying ability. Performances on a particular hands-on science experiment, for example, might be of interest primarily as indicators of students' general abilities to use scientific procedures.

Because performance assessment tasks are often complex and time consuming, students usually cannot be observed on more than a small number of standardised tasks. As a result, questions about:

- the number of tasks required; and
- the extent to which performances can be generalised to other, similiar tasks



Planning performance assessments **Generalisability**

In an attempt to ensure that meaningful conclusions can be drawn about students' levels of attainment from their performances on a limited number of tasks, the Australian Music Examinations Board (AMEB) samples a wide range of student performance.

In Fifth Grade violin, for example, students perform four studies and pieces, one from each of four lists, plus an extra list requirement of three works. Each of the studies and pieces requires students to play in a different style. Examiners look for technique and understanding in each of the styles. Assessment criteria include: tone, intonation, musicianship, phrasing and articulation.

Planning performance assessments

Standardisation

The Third International Mathematics and Science Study (TIMSS) provided training sessions for all task administrators and observers so that the 'circus' of hands-on science tasks was administered in the same manner within and across all countries. This meant that students received the same introductory information and instructions, the same types of assistance when problems were encountered, and had the same time to complete tasks. The Administration Handbook described the information to be provided to students, and provided a suggested introductory script.

may be especially important in high-stakes performance assessments.

Research studies show that in high-stakes performance assessments, students need to be observed on more than one or two performance tasks. Research by Shavelson, Baxter and Xiaohang Gao, for example, suggests that it may be necessary to observe students on up to 15 occasions before an accurate measure of some aspects of science achievement is obtained.[1]

Because complex performance tasks sometimes draw on particular areas of knowledge and skill, careful consideration must be given to the wider context (domain) of performances to which students' observed performances can be generalised. Is it possible, for example, to infer something meaningful about students' levels of physical education achievement from their performances in areas as different as soccer, tennis, basketball, and running?

standardisation

The way in which a performance task is introduced or explained and the degree of assistance students receive in completing the task can affect students' performances. In

Judging performance assessments

Specifying assessment criteria

Assessment criteria for part of the U.K. science task 'Investigating Floating and Sinking' were specified as follows:[2]

- Sees the need to use the kitchen-type scales with a dial because he or she cannot quantify the small difference in weights between any two of the objects using classroom balance scales. Weighs the two objects to the nearest labelled division using kitchen-type scales with a dial.
- Draws a table/chart or contributes towards the construction of a table/chart and accurately records findings for seven out of eight objects.
- Gives a written or oral account of what he or she did which includes each of the stages in the correct order.
- Communicates in written account or orally that he or she knows the difference between a 'fair' and an 'unfair' test.

contexts where comparability matters, it is important that there is as little variation as possible in the demands made on different students, and in the opportunities they have to show their achievement. The greater the requirement for comparability, the more carefully the administration of performance tasks is likely to be controlled or *standardised.*

Assessors may be trained or given standardised administration instructions so that students receive the same introductory information and instructions, the same types of assistance when problems are encountered, and have the same time to carry out the tasks.

judging performance assessments

Performance assessment depends on human judgement. Two people viewing the same performance may recognise different qualities in the performance and so judge the performance differently. Assessors also may by swayed by qualities in a performance that are not central to the skills being assessed. A petite elfin-like gymnast may receive a higher rating than a physically more mature competitor for reasons of physical attractiveness rather than superior performance, for example.

Judgements of performances can be made more comparable by:

- carefully defining the criteria for the assessment of the performance;
- ensuring consistency of criteria interpretation;
- using a number of judges; and
- identifying and taking account of systematic differences among assessors.

specifying assessment criteria

Performance assessment requires judgement of the quality of, and sometimes the process of arriving at, a complex response. Assessment criteria must be well conceived and explicitly defined in an attempt to minimise the influence of irrelevant features of a student's performance or assessors' judgements.

Assessment criteria sometimes are supported by examples of students' performances. Toronto Benchmarks, for example, provide teachers with video-taped examples of speaking performances.

Judging performance assessments ***Multiple judges***

In the Victorian College of the Arts dance assessment, five judges observe each student's performance. Students' performances are scored on a scale of 0 to 100 on the basis of the following criteria:

technical accuracy;
musicality and rhythmic ability;
use of style and presentation;
ability to maintain placement and formation; and
ability to demonstrate an understanding of their own body (kinetic awareness).

To calculate a student's exam mark the highest and lowest scores are deleted. The three remaining marks are averaged. If the result is between two whole numbers the mark is rounded up or down to the nearest whole number.

Teachers can practise using these examples which are accompanied by detailed commentary.

ensuring consistency of interpretation

To make comparable judgements assessors need to develop a shared understanding of the assessment criteria. Providing training sessions in which assessors view and discuss examples of student performance is one method of enhancing comparability. When assessors have the opportunity to work together, greater consistency of criteria interpretation can be achieved through discussion of actual performances.

In some assessment programs, small groups of judges work together over long periods of time, developing a shared 'culture' of assessment.

multiple judges

No matter how clear the assessment criteria, there will always be concerns about the subjectivity of judgements. One response to this concern is to have several assessors judge the same performance and then to average their ratings, possibly ignoring the highest and lowest marks. This method is sometimes used in the assessment of diving, for example.

adjusting for systematic differences among assessors

Another way to improve the comparability of performance assessments is to identify and take into account systematic differences among assessors. Adjustments might be made to take account of a judge's tendency to be unusually harsh, or unusually lenient, when rating performances. This process is sometimes called 'controlling' for differences amoung assessors (in contrast to procedures designed to assure that assessors have similiar understandings of the criteria and assessment process).

Summarising and reporting achievement

Criteria achieved

When U.K. teachers assess student performances using the 'Explanation of Science' Standard Assessment Task, they use criteria drawn from Levels 1 - 3 of the Attainment Targets. To estimate a student's level on this science progress map, teachers use a numerical rule illustrated below.[3]

Level	*Criteria Achieved*			
3			✓✓	✓✓✓✓✓✓
2			✓✓✓	
1	✓	✓✓		
Rule	Fewer than two Level 1 criteria	Two Level 1 criteria	Five or more mixed Level 2 & 3 criteria	Six or more Level 3 criteria
Overall Assessment	Working towards Level 1	Level 1	Level 2	Level 3

summarising and reporting achievement

Judgements on performance tasks can be used to make an overall (on-balance) assessment of a student's level of attainment in a learning area. The procedures for using performance judgements to make overall assessments parallel the procedures available for classroom performance (see pages 19 to 36).

The example at the bottom of page 42 shows how judgements of *criteria achieved* during the performance of a science task are used in one national assessment program to estimate the levels at which students are working.

While students are engaged in the task, teachers mark off criteria as they are demonstrated. These criteria are drawn from several levels of the science attainment tasks. Teachers then use a numerical rule to calculate a student's overall level of achievement.

In some high-stakes performance assessments judges make *holistic* ratings of performances which are then combined across judges to provide an estimate of a student's overall level of achievement.

Sometimes assessors rate different aspects of a student's performance and then combine these *analytic* ratings to obtain an overall (on-balance) measure of achievement. To estimate an 'Exit Level of Achievement' for Dance in one high-stakes end of school program, for example, teachers first rate three aspects of students' Dance performances, 'Choreography', 'Performing', and 'Appreciating', each on a 5-point scale E to A. In rating 'Performing' teachers pay attention to six criteria (see page 44): 'Alignment', 'Body Mechanics', 'Spatial Awareness', 'Rhythmic Awareness', 'Use of Style', and 'Communication in Performance.'

On the basis of these ratings, teachers then obtain a single Exit Level of Achievement from 'Very Limited' to 'Very High' achievement using a numerical process. For example, a student awarded B (4 points) for Choreography, A (5 points) for Performing and B (4 points) for Appreciating would have an aggregate of 13 (4+5+4) - an Exit Level of 'High Achievement'. This numerical process is described in detail on page 45.

Summarising and reporting achievement

Analytic ratings

To estimate 'Exit Levels of Achievement' for students in final year Dance in Queensland, teachers first assess three aspects of dance performances: 'Choreography', 'Performing', and 'Appreciating'. A 5-point scale E to A, is used for each of these three assessments. The scale used to assess Performing is shown below. Teachers then estimate a single Level of Achievement in Dance using the numerical process shown opposite.[4]

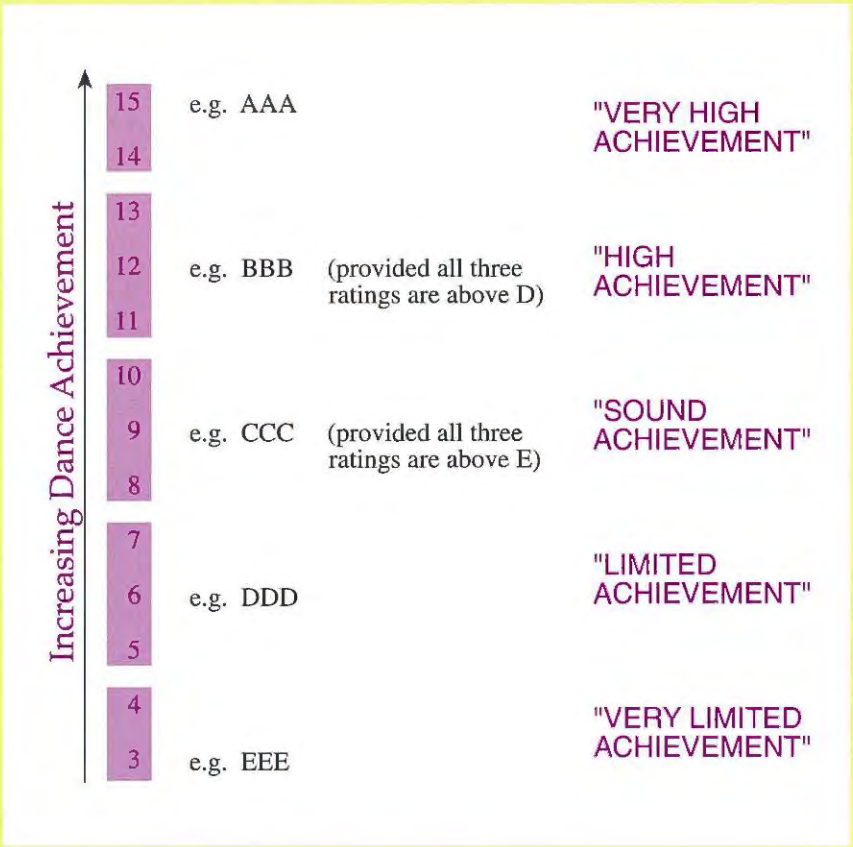
The student consistently demonstrates: **A**	• thorough understanding of alignment in locomotor and non-locomotor movement; • strength, balance, coordination and control in efficient body mechanics; • highly developed spatial awareness; • highly developed rhythmic awareness; • sensitivity in appropriate use of style; • suitable energy and focus in communication with an audience in performance.
The student demonstrates for the most part: **B**	• substantial understanding of alignment in locomotor and non-locomotor movement; • strength, balance, coordination and control in efficient body mechanics; • well developed spatial awareness; • well developed rhythmic awareness; • appropriate use of style; • suitable energy and focus in communication with an audience in performance.
The student demonstrates: **C**	• a variable ability in alignment in locomotor and non-locomotor movement; • some strength, balance, coordination and control in body mechanics; • spatial awareness in simple movement sequences; • rhythmic awareness in simple movement sequences; • some appropriate use of style; • inconsistent energy and focus in communication with an audience in performance.
The student inconsistently demonstrates: **D**	• appropriate body alignment; • use of body mechanics with strength, balance, coordination and control; • use of spatial awareness in most movement sequences; • use of rhythmic awareness in most movement sequences; • use of appropriate style; • communication with an audience in performance.
The student rarely demonstrates ability to: **E**	• align the body; • use body mechanics with strength, balance, coordination and control; • use spatial awareness in most movement sequences; • use rhythmic awareness in most movement sequences; • use appropriate style; • communicate with an audience in performance.

Summarising and reporting achievement

Analytic Ratings

Teachers estimate a single 'Exit Level of Achievement' in final year Dance in Queensland on a scale from, 'Very Limited Achievement' to 'Very High Achievement' using the numerical process shown below. Teachers obtain an aggregate score between 3 and 15 by summing the points awarded in each of three areas of Dance (Choreography, Performing, and Appreciating).

A student can achieve a maximum of 5 points for each area. For example, a student awarded 4 points (B) for Choreography, 5 points (A) for Performing and 4 points (B) for Appreciating has an aggregate of 13 (4+5+4) and thus an Exit Level of 'High Achievement'.



1 Shavelson, R. and Baxter, G. (1992) 'What We've Learned About Assessing Hands on Science', *Educational Leadership,* Vol. 49, No. 8, pp. 20-25 and Shavelson, R., Baxter, G. and Xiaohang Gao (1993) 'Sampling Variability of Performance Assessment', *Journal of Educational Measurement,* Vol. 30, No. 3, pp. 215-232.

2 Department of Education and Science School Examinations Assessment Council (1991) *Assessment Record Booklet, Key Stage 1,* Lincoln: NFER - Nelson, p. 48.

3 Department of Education and Science School Examinations Assessment Council (1991) *Assessment Record Booklet, Key Stage 1,* Lincoln: NFER - Nelson, p. 46.

4 Board of Senior Secondary School Studies, Queensland *Senior Dance* Queensland: Department of Education.

designing performance assessments

In developmental assessment, teachers monitor student progress against a map of developing skills, knowledge and understandings. Performance assessments are one method that can be used to collect evidence of student achievement.

There are many contexts for, and kinds of, performance assessment. Different kinds of performance assessment provide different kinds of evidence that teachers can use to estimate students' levels of attainment.

This article lists issues that teachers need to consider when designing performance assessments. These issues include questions of assessment purpose, methods for judging and recording student performance, and ways to estimate students' locations on a progress map. A 'checklist' summary of the design process is included at the end of the article.

planning performance assessments

what is the purpose of the assessments?

Formative or summative assessment? Are you interested in ongoing informal monitoring for formative or diagnostic purposes, or in developing performance assessment 'events' for summative assessment purposes?

How you answer this question will determine your method of performance assessment. You may choose to use observation and recording strategies like those suggested on pages 9-18 or consider the development of planned assessment events. You may choose to use both strategies, recording anecdotal performance evidence and focusing your observations on specific outcomes with occasional planned tasks.

What curriculum goals or outcomes will you collect evidence about? If you are recording informal observations of student performance:

Do your observations focus on explicit instructional or outcome goals?

Is the evidence you are collecting relevant?

If you are observing performance assessment events:

How well does the assessment measure what it is designed to measure (validity)?

Is there a match between the tasks and the doing outcomes being assessed?

What aspects or details of the performance will you observe? Can the assessment be used to infer performance in other contexts (generalisability)?

How many tasks will you use?

Are the tasks fair to all students?

Do the tasks allow students to show their strengths and weaknesses? (Are they sufficiently open ended?)

In summary:
What is the purpose of the performance assessment?

- formative or summative?
- evidence of which curriculum goals/progress map outcomes?
- is the evidence relevant?
- which details are to be observed?
- which tasks and how many?

focusing and recording observations/ judging performances

if you are informally observing student performances:

Which outcomes will you target and how? Will you look at a range of outcomes or 'spotlight' a small number?

Which students will you observe and how? Will you observe a large number of students or 'spotlight' a few?

Will you use particular techniques such as 'focused questioning' (see p. 11)?

How will you record your observations? What method will you use (e.g. file cards, computer records, annotated class lists, observation grids, 'post-it' notes)?

Is your method efficient?

Does it provide you with readily accessible information for teaching purposes and for feedback to students and parents?

Does it allow you to see growth over time?

if you are observing performance assessment events:

How will you judge the performances? Will you make holistic judgements or use an analytic marking guide?

What assessment criteria will you use? Are they fair? (Have you considered the gender, cultural and language background of the students?)

Are they easily communicated?

If you are developing tasks in a context where the comparability of performances is particularly important you will need to consider how consistently the assessment measures student performance (reliability). You may need to control the conditions of the assessment, ensure consistency of criteria interpretation (marker training and moderation procedures) and use a number of judges (see pages 39-40).

In summary:
How will you focus and record observations and how will you judge performance events?

- Which outcomes will you target and how?
- Which students will you observe and how?
- How will you record your observations?
- How will you judge performances?
- Which criteria will you use?
- How reliable is the assessment?

placing students on a progress map

how will you estimate and report students' locations on a progress map?

By level or location on a continuum? Do you want to place a student 'in a level' and report this level, or position students on the continuum and report this location?

Subjective or objective estimate?
Do you want to make an on-balance subjective estimate or an on-balance statistical estimate of a student's location on a progress map?

In summary:
How will you estimate and report a student's location on a progress map?

- by level or location on the continuum?
- subjective or objective estimate?

A Summary of the Performance Assessment Design Process

Performance task design stage	Task design strategies
Deciding performance assessment purpose	• Describe the assessment purpose. • Review these descriptions against important curriculum objectives and progress map outcomes of the learning area.
Deciding curriculum goals or outcomes to be targeted.	• List the goals or outcomes. • Review these against important curriculum objectives and progress map outcomes of the learning area to check relevance.
Deciding details of performances to be observed.	• Decide the number of performance 'events'. • Review for open-endedness. • Review for generalisability. • Review for efficiency, accessibility of information, ability to show growth over time. • Review for fairness (gender, culture, language).
Deciding a method for focusing and recording observations and judging performance tasks.	• Decide how you will focus on outcomes. • Decide which students you will observe. • Decide how you will record your observations. • Decide how you will judge performances (assessment criteria). • Review for easy communication.
Deciding a method for estimating and reporting locations on a progress map.	• Describe the method for estimating locations on a progress map. • Describe the method for reporting locations on a progress map. • Review these descriptions against the task purpose and audience.

Please photocopy and fax or mail.

ASSESSMENT RESOURCE KIT (ARK)

ORDER FORM

Australian Council for Educational Research
19 Prospect Hill Road (Private Bag 55)
Camberwell Vic 3124 AUSTRALIA
Telephone: (03) 9277 5656 Fax: (03) 9277 5678
International: Tel: 61 3 9277 5656 Fax: 61 3 9277 5678
email: sales@acer.edu.au

Charge to

Name *or*
Organisation _____

Purchase Order no. _____

Address _____

_____

_____ Postcode _____

Deliver to

Name _____

Organisation _____

Street address _____

_____ Postcode _____

Telephone ( ) _____ Fax ( ) _____

Order date _____ Date required _____ ACER account no. _____

Title	Cat. No.	Price	Quantity	Total Price
Developmental Assessment	100ARK	AUD$12.95		
Portfolios	101ARK	AUD$12.95		
Performances	102ARK	AUD$12.95		
Projects	103ARK	AUD$12.95		
Progress Maps	105ARK	AUD$12.95		
Understanding Developmental Assessment (videotape)	700ARK	AUD$49.95		
Set of 3 Posters: Reading, Spelling, Writing	400ARK	AUD$6.95		
Products	104ARK	AUD$12.95		
Forthcoming titles:				
Assessment Methods (available late 1998)	106ARK	AUD$12.95		
Paper and Pen (available late 1998)	107ARK	AUD$12.95		
Implementing Developmental Assessment: Workshop Manual & Videotape (available late 1998)	500ARK	TBA		

ADD FREIGHT:
Within Australia: 10% of invoice value, min $5.00 max $15.00.
Overseas(airmail): 1 - 5 copies: AUD$30
6 - 10 copies: AUD$45
More than 10 copies: please ring or fax for a quote.

NOTE: Prepayment required for non-account orders.
Prices subject to change without notice - you are welcome to check before you order.

Subtotal _____
Freight _____
TOTAL $ _____

CHARGE: ☐ ACER Account No. _____ ☐ Cheque Enclosed (AUD$)

☐ Bankcard ☐ Mastercard ☐ American Express ☐ Visa

ACER PRESS

☐☐☐☐ ☐☐☐☐ ☐☐☐☐ ☐☐☐☐

Name (*please print*) _____ Signature _____ Expiry date _____

ACN 004 398 145

Performances is one in a series of magazines in the ACER Assessment Resource Kit (ARK).

This video and magazine resource provides information about assessment issues and methods.

For further details about other magazines, videos and the workshop manual in this series contact the Australian Council for Educational Research, 19 Prospect Hill Road, Camberwell, Victoria, Australia, 3124.
Phone: +61 3 9277 5656
Facsimile: +61 3 9277 5678

ACER