Developmental Assessment

ASSESSMENT RESOURCE KIT

Portfolios
Performances
Projects
Products
Paper & Pen

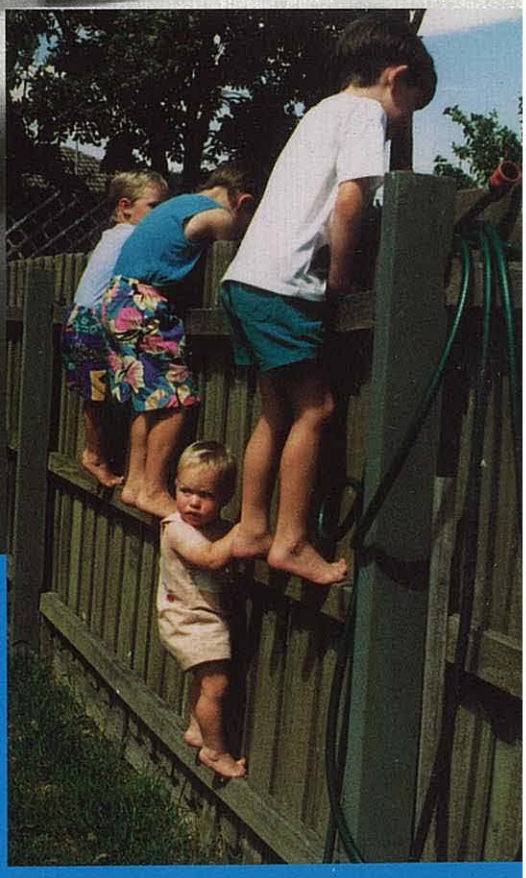
Progress Maps

Assessment Methods

Judging and Recording

Estimating Attainment

Reporting



ISBN 0-86431-204-0

ARK

Geoff Masters & Margaret Forster

Contents

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

Reprinted 1996, 1997

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

ISBN 0 86431 204 0

Edited by Eve Recht
Designed by Benjamin Forster

Printed by Allanby Press Printers Pty. Ltd.

Photographs courtesy of Village School, North Croydon, Knox Grammar, Wantirna South, Luther College, Croydon, Kingswood Primary, Outer Eastern College of TAFE and Waldau Primary, Doncaster.

Thanks to teachers in the National Schools Network for working with ACER in the trial of these materials (Burwood Girls High School, NSW; Ardrossan Area School, SA; Charles Condor Primary, ACT; Yea High School, VIC; and St Bernadette's Primary School The Basin, VIC).

what is developmental assessment?

Developmental assessment is the process of monitoring a student's progress through an area of learning so that decisions can be made about the best ways to facilitate further learning.
The unique feature of developmental assessment is its use of a progress map (or 'continuum'). A progress map describes the nature of development—or progress or growth—in an area of learning and so provides a frame of reference for monitoring individual development.

In developmental assessment, progress is monitored in much the same way as a child's physical growth is monitored: from time to time an estimate is made of a student's location on a developmental continuum, and changes in location provide measures of growth over time.

progress maps

A progress map is a picture of the path students typically follow as they learn.

This vertical map provides a description of skills, understandings and knowledge in the sequence in which they typically develop: a picture of what it means to 'improve' in an area of learning.

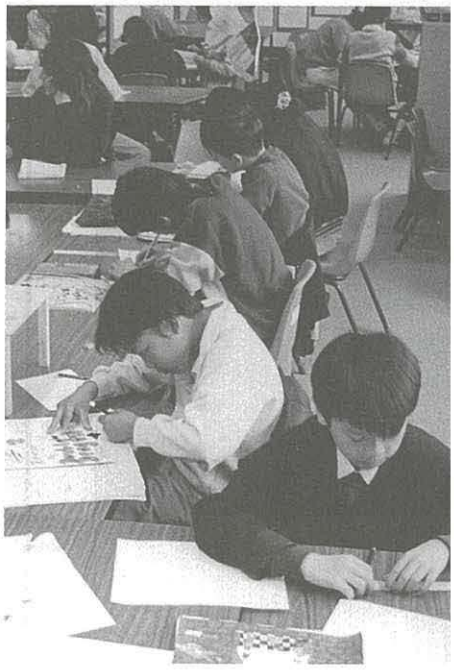
The first step in implementing developmental assessment is to construct a progress map.
A progress map usually begins with teachers' understandings. Through their day-to-day experiences, teachers gain an understanding of how student development usually occurs in an area of learning. They come to recognise indicators of progress. An initial sketch of a progress map is made by putting on paper the understandings that teachers already have.

This initial sketch is then tested. Do other teachers agree with this description of typical growth? What is the empirical evidence for this map as a picture of typical student progress? Is this picture consistent with theoretical understandings of how learning occurs? How useful is the resulting progress map in practice? In other words, does it provide teachers with a useful framework for thinking about and monitoring student development through an area of learning?

The order of learning outcomes on a progress map sometimes reflects a natural, perhaps inevitable, developmental order. All children probably develop an understanding that spoken language can be represented using marks on paper before they understand the meanings of particular written words, for example. Knowing how to add and subtract integers may be a prerequisite to learning how to add and subtract fractions or decimals. In other cases, the order in which skills, understandings and knowledge develop is influenced by common classroom conventions: a decision by teachers to teach some skills and knowledge ahead of others.

The intention in developmental assessment is to obtain an estimate of a student's current location on a progress map as a guide to the kinds of learning experiences

The developing skills, understandings and knowledge described on a progress map go by a variety of names, including 'indicators', 'descriptors', and 'outcomes'.



5

Uses unitary ratios of the form 1 part to X parts
(The ratio of cordial to water was 1 to 4)
Understands that common fractions are used to describe ratios of parts to whole (2 in 5 students ride to school. In school of 550, 220 ride bikes)
Uses percentages to make straightforward comparisons
(26 balls from 50 tries is 52%; 24 from 40 tries is 60%, so that is better)
Uses common equivalences between decimals, fractions and percentages
('One-third off is better than 30% discount')
Uses whole number powers and square roots in describing things
(finds length of side of square of area 225 sq cm as square root of 225)

4

Counts in decimal fraction amounts ('0.3, 0.6, 0.9, 1.2, ... ')
Compares and orders decimal fractions
(orders given weight data for babies to two decimal places)
Uses place value to explain the order of decimal fractions
(which library book comes first 65.6 or 65.126? why?)
Reads scales calibrated in multiples of ten
(reads 3.97 on a tape measure marked in hundredths, labelled in tenths)
Uses the symbols =, < and > to order numbers and make comparisons
(6.75 < 6.9; 5 x $6 > 5 x $5.95)
Compares and orders fractions (one-quarter is less than three-eighths)

3

Counts in common fractional amounts
('two and one-third, two and two-thirds, three, three and one-third')
Uses decimal notation to two places (uses 1.25m for 1m 25cm; $3.05 for three $1 coins and one 5c coin; 1.75kg for 1750g)
Regroups money to fewest possible notes and coins
(11x $5 + 17x $2 + 8 x $1 regrouped as 1x $50 + 2x $20 + $5 + $2)
Uses materials and diagrams to represent fractional amounts
(folds tape into five equal parts, shades 3 parts to show 3/5)
Expresses generalisations about fractional numbers symbolically
('1 quarter = 2 eighths' and '1/4 = 2/8')

2

Counts forwards and backwards from any whole number, including skip counting in 2s, 3s and 10s
Uses place value to distinguish and order whole numbers
(writes four ten dollar notes and three one dollar coins as $43)
Estimates the size of a collection (up to about 20)
Uses fractional language (one-half, third, quarter, fifth, tenth) appropriately in describing and comparing things
Shows and compares unit fractions (finds a third of a cup of sugar)
Describes and records simple fractional equivalents ('The left over half pizza was as much as two quarters put together')

1

Counts collections of objects to answer the question 'How many are there?'
Makes or draws collections of a given size
(responds correctly to 'Give me 6 bears')
Makes sensible estimates of the size of small collections up to 10
(for 7 buttons, 2 or 15 would not be a sensible estimate, but 5 would be)
Skip counts in 2s or 3s using a number line, hundred chart, or mental counting ('2, 4, 6 ... ')
Uses numbers to decide which is bigger, smaller, same size
(If he has 7 mice at home and I have 5, then he has more)
Uses the terms first, second, third ('I finished my lunch second')

The lower portion of a Counting and Ordering progress map (Australian mathematics profile) showing examples of knowledge, skills and understandings.[1]

likely to be most useful at that stage in the student's learning and as a basis for monitoring growth over time.

This feature of developmental assessment distinguishes it from other forms of assessment which do not monitor progress against described continua. If the purpose of an assessment activity is to establish whether or not a student has satisfactorily completed a set of assigned tasks, if the purpose is only to rank students so that some can be admitted to an educational course or be offered a scholarship, or if the result of an assessment is an uninterpreted score or grade, then that does not constitute developmental assessment.

assessment methods

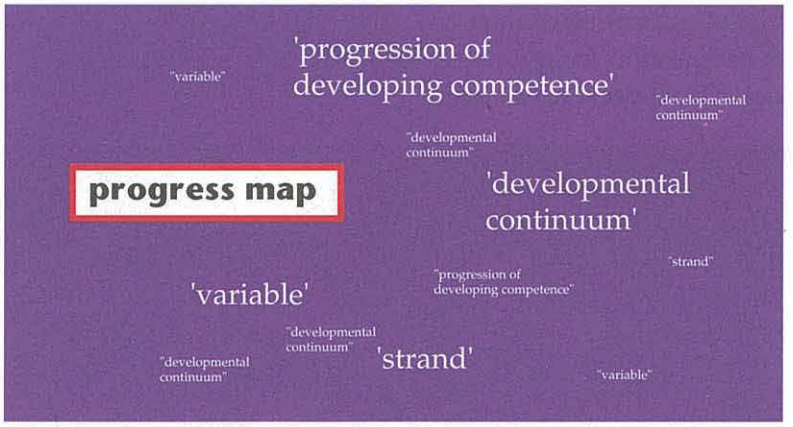
The second step in implementing developmental assessment is to collect evidence which can be used to estimate students' locations on a progress map.

In developmental assessment, records of observations provide the evidence needed to estimate students' levels of attainment (i.e. locations) on a map. In general, the larger the number of relevant observations, the richer the available evidence and the more dependable the conclusion about a student's current level of attainment.

A requirement of observations is that they be *relevant*. They must provide evidence about the area of learning to be assessed. More exactly, they must provide evidence about the learning outcomes identified on a progress map. Unless observations are relevant, conclusions based on those observations lack validity.

Some learning outcomes require observations of a particular kind. Skills in designing and undertaking an investigation, collecting, analysing and evaluating information, and writing a report may be best assessed in the context of an assigned *project*. The ability to write for a range of audiences and purposes may be best assessed by observing a collection (*portfolio*) of a student's writing assembled over a period of time. Skills in manipulating apparatus, operating machinery, playing musical instruments, co-operating with others in a team activity, and speaking a second language may be best assessed by observing *performances*. Skills in making items of food, ceramics, textiles, wood and metal may be

The distinguishing feature of developmental assessment is the intention to monitor student progress against a pre-constructed map of developing skills, knowledge and understanding.



Progress maps are known by a variety of names. Some of the more common names are shown here.

best assessed by observing the *products* of student work. And a student's mastery of a body of knowledge, ability to reason logically about a range of problems, and ability to apply procedures such as mathematical algorithms may be most efficiently assessed through *paper and pen* exercises.

Developmental assessment uses all these and other methods of observation. Each method can be used informally as part of a teacher's day-to-day practice or, where high levels of reliability and comparability are important, as a basis for developing more structured assessment tasks and activities.

In collecting evidence it is especially important that observations are not limited to students' performances on passive, reproductive tasks. Students also must be observed creating their own solutions to problems. And the contexts in which they are observed should, wherever possible, be meaningful to students and interesting in their own right.

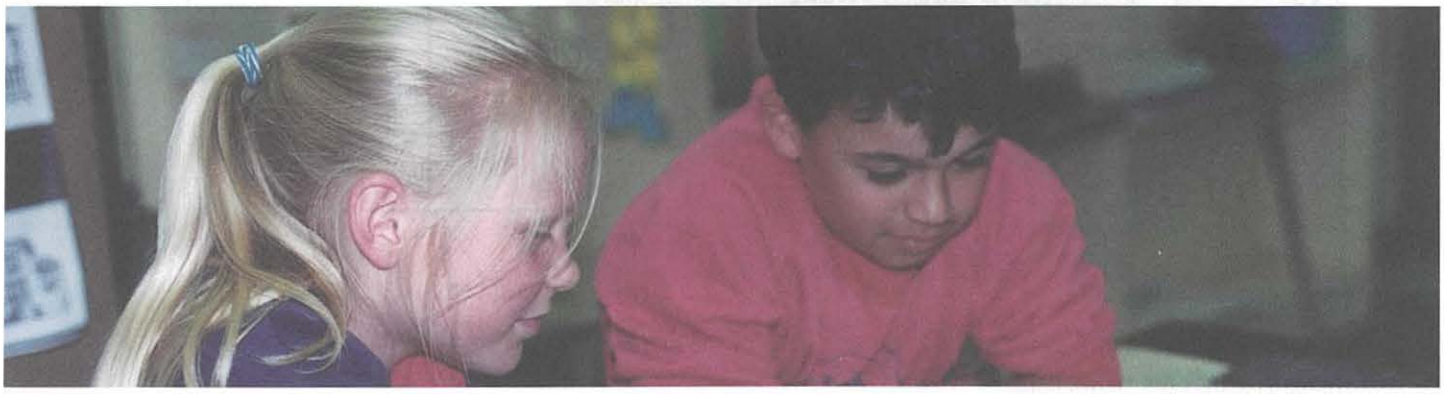
Not all methods of assessment are equally accessible and fair to all students. Teachers need to be aware of the implications of using particular methods of assessment for some individuals. Students may be disadvantaged on an oral presentation because of their cultural background, for example. Students from non-English speaking backgrounds may be better able to demonstrate their skills and knowledge on short-answer paper and pencil questions than on extended writing or speaking tasks. And, relative to boys, girls are likely to perform better on essays and project work than on multiple-choice tests in the same learning area.

judging and recording

In developmental assessment, the collection of evidence includes the systematic recording of observations and judgements. Records of observations and judgements can take many forms.

Teachers use a variety of schemes to record their day-to-day observations of student classroom behaviour. These 'anecdotal' records, which some teachers later transfer to individual student record sheets, provide valuable information which can go unnoticed or be forgotten if not systematically recorded as it occurs.

Teachers make other observations through assigned classroom work. Writing tasks, projects, presentations to the class, portfolio entries, classroom exercises, and tests all provide opportunities to collect information about a student's current level of attainment in an area of learning.



Some of the records teachers keep simply indicate whether or not assigned tasks have been completed correctly (e.g. did the child give a sensible estimate of the number of buttons on the table?). In other cases, more detailed records are kept of the steps a student takes to solve a problem or of students' partially correct but incomplete understandings.

In still other cases teachers record ratings of student work, making separate judgements of different aspects of a piece of work ('analytic' ratings) or single overall ('holistic') ratings. Comments, evaluations and suggestions for improvement are commonly recorded directly on to work as feedback to students.

When making judgements of student work, teachers need to be aware of common observational 'errors'. Examples of such errors include the tendency of observers to see what they expect to see on the basis of their preconceptions of a student's ability, and the tendency of observers to under-rate the performances of students of a particular sex or cultural background on some kinds of tasks and in some contexts.

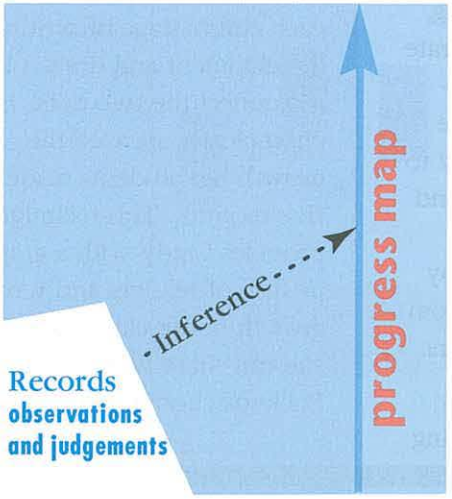
estimating attainment

The third step in developmental assessment is to use the available evidence (records of observations) to draw a conclusion (or inference) about an individual's current location on a progress map.

Because a progress map is a description of the path of typical student progress through a learning area, the record of observations made for any particular student will only more or less resemble this path of typical progress. Developmental assessment requires an 'on-balance' estimate of a student's location on a progress map. Three questions can be asked about this estimate:

- How valid is it?
- How reliable is it?
- How objective is it?

Records of observations and judgements can take many forms including anecdotal notes, records of performances on tasks and ratings of student work.



Developmental assessment requires an on-balance decision (inference) about a student's location on a progress map based on the available evidence.

The validity of an estimate depends on the relevance of the observations on which it is based.

The reliability of an estimate depends on the amount of information on which it is based.

The objectivity of an estimate depends on the extent to which it is unaffected by *who* makes the observations and *what* exactly students are observed doing.

The *validity* of an estimate depends on the relevance of the observations on which it is based. Do the observations provide evidence about the full range of outcomes that make up the learning area? Is the available evidence an adequate reflection of the student's abilities, or is it too dependent on other influencing factors such as competence and confidence in the use of English? Are the observations 'fair' indicators of achievement in the sense that they are not affected by irrelevant student characteristics such as cultural background or gender?

The *reliability* of an estimate depends on the amount of information on which it is based. In general, the greater the amount of evidence used in making an estimate, the more reliable that estimate.

The *objectivity* of an estimate depends on the extent to which it is unaffected by the choice of tasks or the choice of assessors.

In day-to-day work students are not always observed in the same contexts or undertaking the same tasks. Students who are able to apply skills and knowledge to more complex, more demanding, and less familiar tasks demonstrate higher levels of ability than students who are able to apply the same skills and knowledge only to less complex, less demanding, and more familiar tasks. Different students also may be assessed by different assessors, some of whom may be more lenient than others.

Objective estimates sometimes require procedures for identifying and taking account of differences in task difficulty and assessor harshness.

reporting

In developmental assessment, student achievement is monitored and reported in terms of progress maps (or developmental continua). The monitoring of achievement against an explicit map or continuum permits a variety of graphical displays and descriptive interpretations of student progress.

A progress map provides a framework for monitoring student growth in an area of learning. Because this growth is monitored against a described continuum, individuals' estimated locations on a progress map can be interpreted and reported descriptively in terms of the skills, knowledge and understandings typically demonstrated by students at those locations.

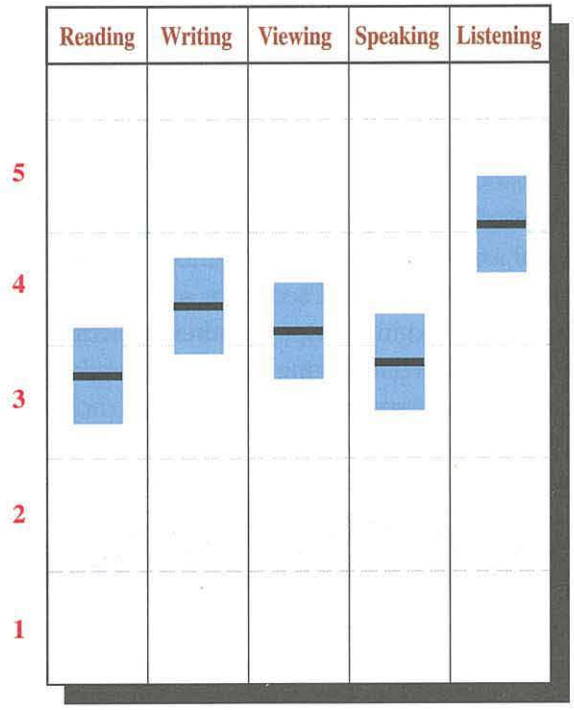
At the end of the first year of using the continuum, Cindy Ruptic graphed the reading and writing development of her first and second graders on a chart. She used a dot on the left to indicate each child's stage of writing development in the fall. In the spring, she made a second dot to indicate each child's stage of writing development and drew a line to connect the two dots. Her chart clearly showed the growth her students made in five months. This technique provides Cindy with a graphic picture of reading and writing growth in her classroom that she can share with parents on Back-to-School night.

Bonnie Campbell Hill[2]

These descriptions of student achievement can then be used as a focus for teacher-student and teacher-parent discussions of progress and to identify with students and parents the kinds of learning activities likely to be most useful at particular stages in a student's learning. In some classrooms, teachers make displays of progress maps and place them on classroom walls for students to see and discuss. In some schools and education systems, parents are given materials explaining key progress maps and receive reports of student progress in relation to those maps.

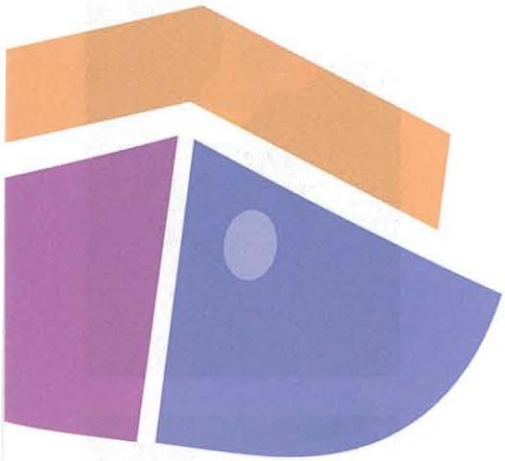
Once a progress map is constructed, it provides a framework for thinking about levels of educational achievement and a basis for setting goals for improvement. What level of attainment in Reading is it reasonable to expect most Year 3 students to reach by the end of the school year? What level of attainment in Measurement is it reasonable to expect most Year 6 students to reach by the end of the year? Questions of this kind lead to the development of *standards*: expectations of student achievement by particular stages in the schooling process.

As well as providing a framework for monitoring and reporting the progress of individual students, a progress map also provides a frame of reference for monitoring and reporting the progress of groups of students. These groups may be individual classes, all students in a school, all students in an education system, or particular subgroups of a student population such as girls, boys, or students from non-English speaking backgrounds.



Display of one student's estimated levels of achievement on five language proficiency progress maps.

A 'standard' is a desired level of attainment.



Developmental Assessment

A focus on outcomes

Developmental assessment shifts the focus in assessment from notions of 'passing' and 'failing' to the concept of ongoing growth: from an emphasis on comparing one individual with another to an emphasis on students' developing skills, knowledge and understandings.

A progress map

Developmental assessment provides a broader perspective on student growth than is available in a single classroom. A progress map draws on experience and evidence from a wide range of classrooms. It also sets student growth in the context of progress made in earlier years of school and progress that can reasonably be expected in the future. In other words, a progress map provides a 'whole-school' view of learning.

A variety of evidence

Developmental assessment uses a rich variety of evidence about a student's achievements in an area of learning. The range of skills, knowledge and understandings that make up most learning areas require a variety of assessment methods. For many learning outcomes, the most appropriate methods of assessment involve teacher observation and judgement of students' performances and work.

On-balance decisions

Developmental assessment requires on-balance decisions (inferences) about students' locations on a progress map on the basis of available evidence. Because students demonstrate skills, knowledge and understandings from a range of levels along a progress map, it is necessary to weigh the available evidence to make a 'best' estimate of an individual's current level of attainment.

Graphical and descriptive reporting

Developmental assessment supports methods of reporting which are more informative than raw test scores or scores that show only where individuals stand in relation to other students. A variety of graphical and descriptive reporting methods have been proposed for use in developmental assessment.

1 Curriculum Corporation (1994) *Mathematics Profile for Australian Schools*, Carlton: Curriculum Corporation, pp. 26, 40, 56, 70 & 86.

2 Campbell Hill, B. & Ruptic, C. (1994) *Practical Aspects of Authentic Assessment: Putting the Pieces Together*, Norwood M.A.: Christopher-Gordon Publishers Inc., p. 242.

progress maps

A central feature of developmental assessment is its use of a progress map as a frame of reference for assessing and reporting student learning.

A progress map—also sometimes known as a 'developmental continuum', 'progression of learning', or 'variable'—describes the path of typical student progress through an area of learning. It is a shared framework for monitoring student growth.

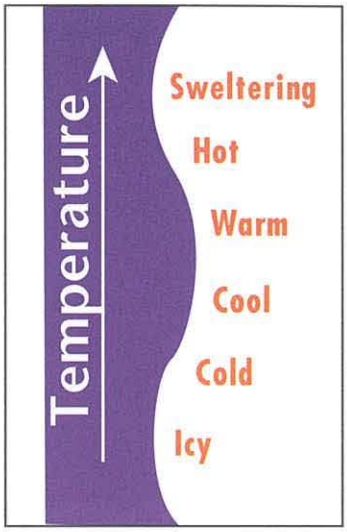
The concept of a progress map or continuum as a shared communication device is a familiar idea when dealing with physical attributes such as height, weight or temperature. In the case of temperature, for example, it is usual to visualise a temperature continuum extending from very low temperatures at one extreme to increasingly high temperatures at the other. Various numerical systems (Celsius, Fahrenheit, Kelvin) have been developed as alternative ways of referring to positions on this continuum. But it also is common in everyday conversation to describe temperatures in terms of broad and fuzzily-defined temperature ranges such as 'warm', 'hot' and 'sweltering'.

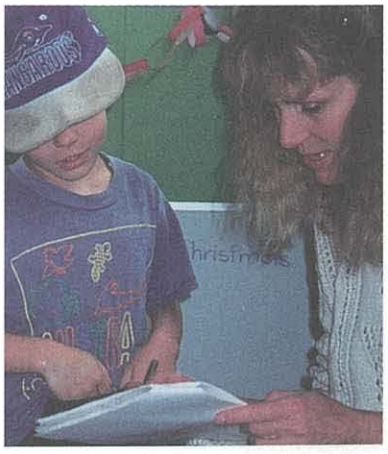
It is possible to imagine observations that might be made on 'hot' days, other observations that might be made on 'warm' days, and still other observations that might be made when the temperature is 'sweltering'. These possible observations could be written down as descriptions of temperature levels. Together they would provide a descriptive temperature 'map' and begin to give everyday meaning to the temperature continuum. These descriptions also could be used as pointers to, or indicators of, the temperature on any particular day.

In a similar way it is common to visualise student progress in an area of learning as occurring along a continuum from lower-level, rudimentary skills and understandings to higher-level, more sophisticated skills and understandings. Indeed, the idea of a continuum of increasing competence is invoked whenever we use such words as 'improving', 'growing' or 'developing'.

There are many examples of progress maps constructed as frameworks for assessing and monitoring student progress in particular areas of school learning. The First Steps program in Western Australia has constructed four progress maps (known as 'developmental continua') in Reading, Writing, Spelling, and Oral Language. These progress maps 'make explicit some of the indicators, or descriptors of behaviour' in these four areas of language development.[1]

The First Steps *Spelling* Continuum is divided into five broad levels labelled 'Preliminary' to 'Independent'. Teachers use First Steps indicators to guide their gathering of information about students' spelling accomplishments through the observation of students as they write and through the inspection of samples of student writing. This information is then used to 'place' students on the Spelling Continuum: that is, to make judgements about individuals' current locations on this progress map.





Some descriptions of progress maps...

- a description of the progression of learning typically achieved by students

  Australian profiles[2]

- a visual representation of development using descriptors to depict the developmental stages of learning

  Bainbridge Island Continua[3]

- an explicit way of mapping children's progress through observation

  First Steps Developmental Continua[4]

- a collection of references for evaluating the growth of individual students

  Toronto Benchmarks[5]

- a progression of increasing competence

  Robert Glaser, 1963[6]

profiles for Australian schools

The concept of a progress map is a fundamental idea

underlying the profiles for Australian schools.

Within each of eight key learning areas—English, mathematics, science, studies of society and environment, technology, the arts, health and physical education, and languages other than English—profiles provide a set of achievement progressions known as *strands*.

In mathematics, for example, the strands are:

- number;
- measurement;
- space;
- chance and data;
- algebra; and
- working mathematically.

These are major groupings of mathematical content, processes and conceptual understandings.

Each profile strand is divided into eight levels. These levels, which are intended to span the compulsory years of schooling (Years 1 to 10), are broad ranges of achievement which together indicate progression in student learning.

Although each child is a unique individual, an overall pattern of growth and development can be traced, within which individual differences can be accommodated. Just as the milestones or indicators of physical growth can be charted, so too can indicators of language and literacy development.

First Steps[7]

Rather than viewing reading and writing as something to be 'taught' in first grade, we now know that children acquire many concepts about print at an early age. Many two-year-olds can spot favorite restaurants and preschoolers often act as though they are 'reading' the books they have memorized. Literacy learning, and indeed all learning, occurs on a continuum.

Campbell Hill and Ruptic [8]

The skills, knowledge and understandings that develop in a learning area are known in the profiles as *outcomes*. Student progress on a profile strand occurs as development in relation to these outcomes.

Because an outcome is a general skill, set of skills, or understanding, there are many possible situations in which an outcome can be applied and demonstrated. Examples (manifestations) of an outcome are known in the profiles as *pointers*. The profiles provide a small number of pointers to illustrate each outcome.

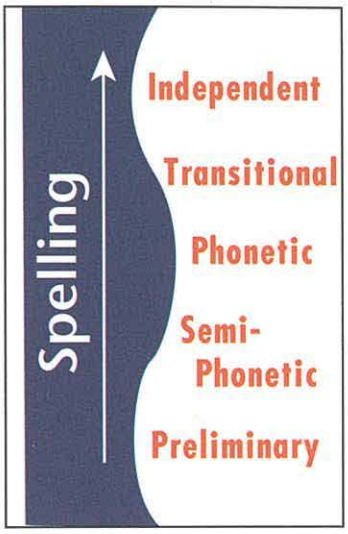
constructing a progress map

Although every progress map is developed with the same intention—to provide a picture of the path of student progress through a learning area as a framework for assessing and monitoring student growth—there are several alternative starting points in the construction of a progress map.

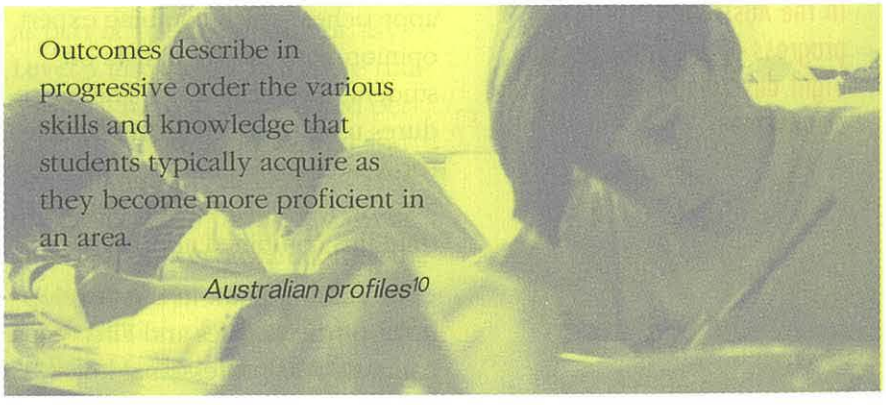
At one extreme it is possible to ask a group of people familiar with an area of learning, perhaps classroom teachers and subject matter specialists, to provide a picture of developing competence as they understand it. To develop this picture teachers and subject matter specialists draw on their knowledge of the learning area and their familiarity with typical patterns of student growth and common classroom conventions.

Some progress maps are based almost entirely on expert opinion. Examples of this 'top-down' approach are the 'attainment targets' developed for the national curriculum in England and Wales. Those progress maps, divided into ten levels spanning ages 7 to 16, were developed to relate to 'expected routes of educational development giving some continuity to a pupil's assessment at different ages'.[9] Each progress map provides a 'framework of progression' based primarily on professional opinion.

At the other extreme a progress map can be constructed from an analysis of how students perform on assigned activities and tasks. This extreme can be described as a 'bottom-up' approach because the resulting progress map is based not on expert opinion, but on actual student performances.



The First Steps Spelling Continuum is divided into five levels known as 'phases'.



Outcomes describe in progressive order the various skills and knowledge that students typically acquire as they become more proficient in an area.

Australian profiles[10]

Outcome and Examples

Mathematics Outcome (from Number strand, Level 3)

Estimates and calculates mentally, including adding (sum to 100) and subtracting two-digit numbers and multiplying numbers to 10.

Examples of Outcome

- '3 eights is 24, so 6 eights would be double that, so 6 x 8 = 48'
- '5 eights is 40, so 8 goes into 40 five times.'
- '3 lots of 5 is 15, 3 lots of 50 is 3 lots of five tens, 15 tens, 150.'
- '46 and 23 is 46, 56, 66 plus 3 which is 69.'
- '7 nines is 7 tens take 7 ones, which is 70 take 7 or 63.'
- 'The mugs cost $4.95, so $30 will be enough for 6.'

Mathematics Profile[11]

In the Australian profiles, progress occurs through eight described levels of achievement.

Examples of empirically-based progress maps include the Toronto Benchmarks and the scales used for the US National Assessment of Educational Progress.

The Toronto progress maps are 'drawn from students' performances'. They describe a sequence of increasingly sophisticated skills and understandings based on observations of students' performances on assigned tasks. The progress maps developed as reporting frameworks for the US National Assessment of Educational Progress similarly are based on an analysis of how students perform on assigned assessment tasks.

Between these two extremes are approaches which combine expert opinion with empirical data on students' performances. The procedures used to construct the Profiles for Australian Schools and the First Steps developmental continua illustrate this combined approach.

The starting point in the development of the profiles and First Steps continua was professional knowledge and opinion and relevant research evidence. But the first versions of these progress maps were seen only as drafts requiring empirical verification and testing. It was recognised that there would be a need to verify and perhaps change the order of descriptors (outcomes/indicators) along these progress maps on the basis of data collected in classrooms. There also would be a need for ongoing revision and enrichment of the progress maps through the addition of other examples and samples of student work.

verifying a progress map

When the starting point in the construction of a progress map is professional knowledge and opinion, it is necessary to test the resulting map against experience. Is the order of skills, knowledge and understandings on the map the order of development as observed and understood by classroom teachers? Is the order consistent with typical observations of student performance?

Those developing the UK attainment targets [progress maps] were not basing them on empirical data collected on the performance of children but more on their experience and beliefs about what children *should* be able to do.

Chris Whetton[12]

Each strand of the Australian profiles in each of the eight learning areas, and each of the First Steps developmental continua was drafted on the basis of expert knowledge and opinion and then tested in classrooms. Teachers were asked to reflect on the attainments of particular students in their classes. Each teacher was given a number of profile outcomes and asked to indicate whether a student demonstrated that behaviour 'hardly ever', 'sometimes', or 'almost always'.

On the basis of these teacher assessments, outcomes were then positioned ('calibrated') on the profile strand. Did 'Level 1' outcomes cluster towards the lower end of the strand? Did 'Level 2' outcomes cluster above 'Level 1'? Were some 'Level 3' outcomes as difficult in practice as some 'Level 4' outcomes? The calibrated order of outcomes along each profile strand provided a picture of development based not on expert opinion, but on teachers' judgements of the achievements of particular students (see page 14).

For some strands, this empirical order matched the preliminary order very closely. For other strands, the match was poor and it was necessary to re-position or re-describe some outcomes to make their meaning clearer.

The process of verifying a progress map is similar to taking a road map drawn from memory and checking its accuracy against the terrain it is intended to represent.

adding detail to a progress map

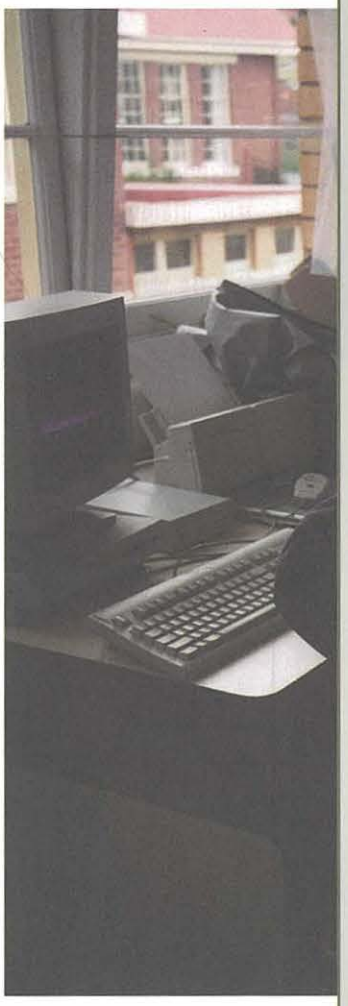
A progress map is never finalised.

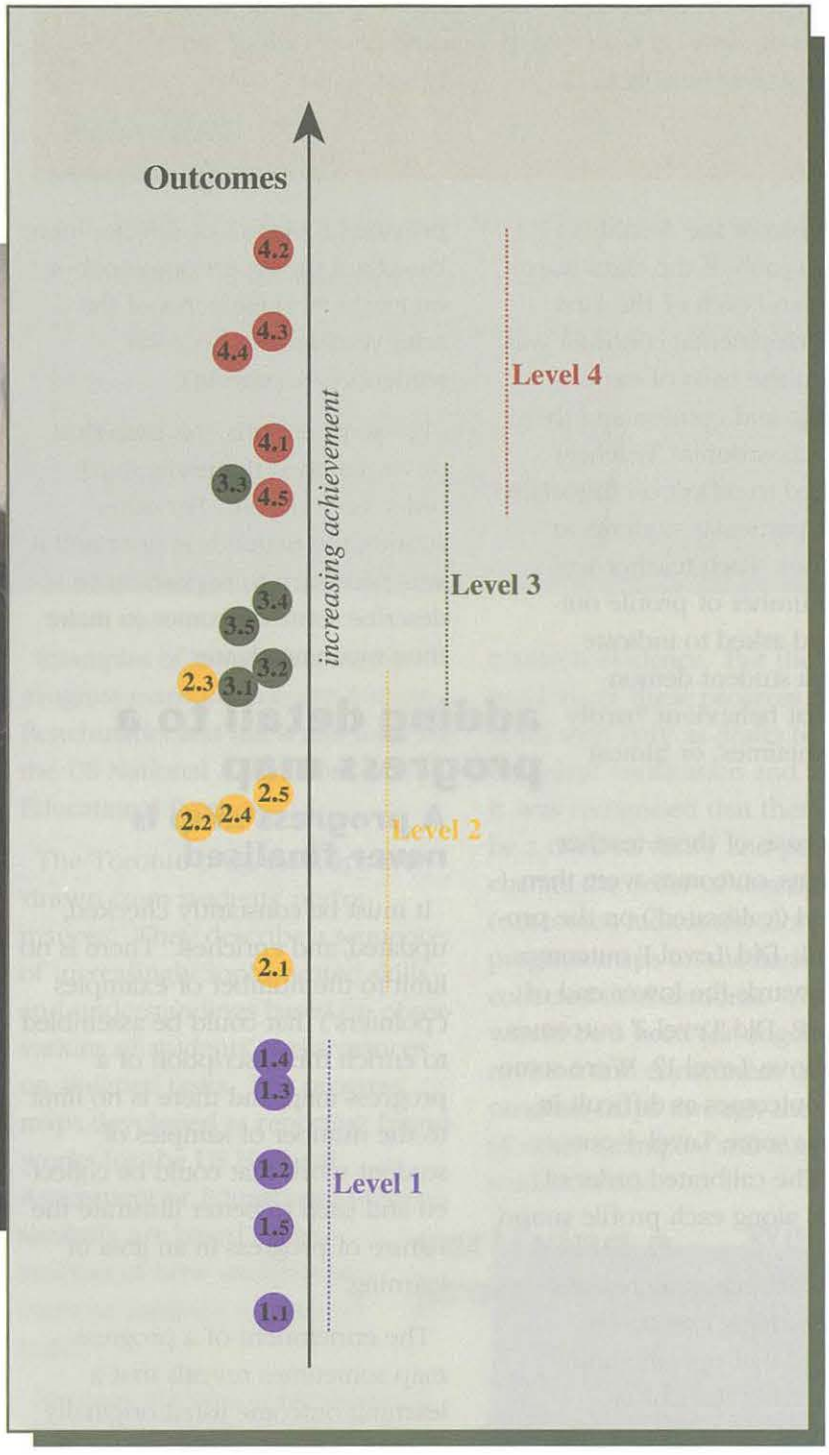
It must be constantly checked, updated, and enriched. There is no limit to the number of examples ('pointers') that could be assembled to enrich the description of a progress map, and there is no limit to the number of samples of student work that could be collected and used to better illustrate the nature of progress in an area of learning.

The enrichment of a progress map sometimes reveals that a learning outcome listed originally as part of the description of, say, Level 3, also occurs in a more rudimentary form at Level 2, and perhaps in a more sophisticated and demanding form at Level 4. This observation can be used to revise and sharpen the description of the outcome at Level 3 and to enrich the definition of Levels 2 and 4 by adding descriptions of that outcome as it occurs at those levels.

The Benchmarks represent what students *can* do on activities that operationalize the Ministry curriculum objectives. This approach to defining standards of student achievement is subtly different from approaches that attempt to define what students *should* do or to state the critical learning outcomes that students are expected to demonstrate at a given point in time.

Toronto Benchmarks[13]





Twenty outcomes from the Space strand of the Australian mathematics profile calibrated on the basis of teachers' judgements of the extent to which individual students had achieved each outcome.

Refining a progress map

The use of student performances to inform and enrich a progress map is illustrated in the DART *Developmental Assessment Resource for Teachers*.[14] DART provides activities and tasks that teachers can use to assist them to place students on the Viewing strand of the Australian profile in English. As part of DART viewing activities, students watch a film and answer a series of questions.

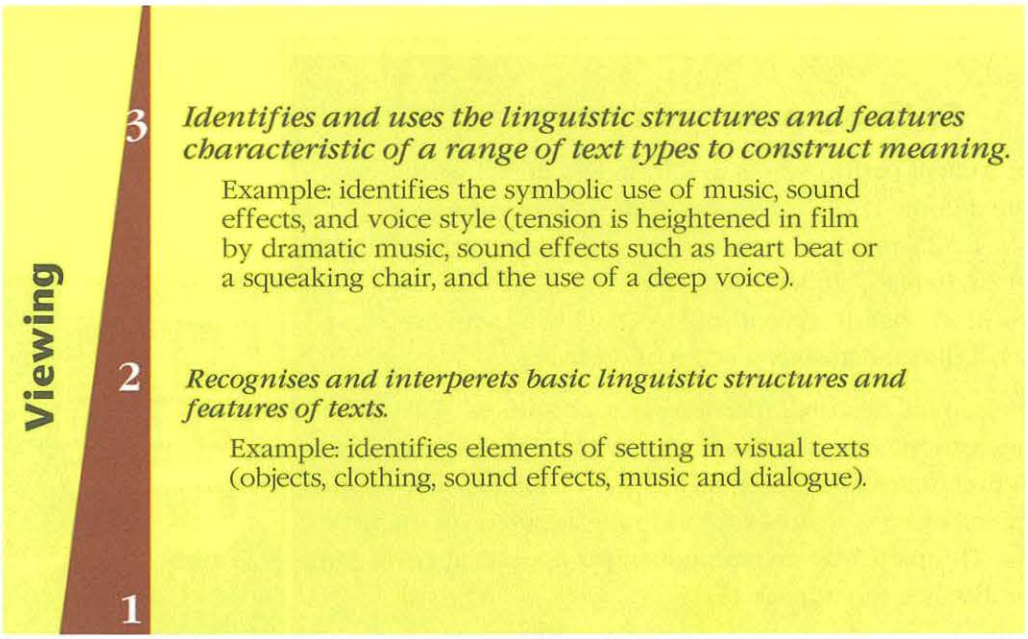
The Viewing strand describes increasingly sophisticated skills and understandings associated with viewed texts such as films and posters. Development in this area of learning includes, for example, a developing understanding of how devices such as music are used for dramatic effect in films. This particular understanding first appears at Level 3 in the English profile (see top of page 16).

Some DART Viewing tasks explore students' understandings of these devices (see Questions 16 and 17 on page 16). Do students understand the purpose of zooming in on eggs in a frying pan at a particular point in the film? Do they understand that music is used for dramatic effect at a key stage of the film?

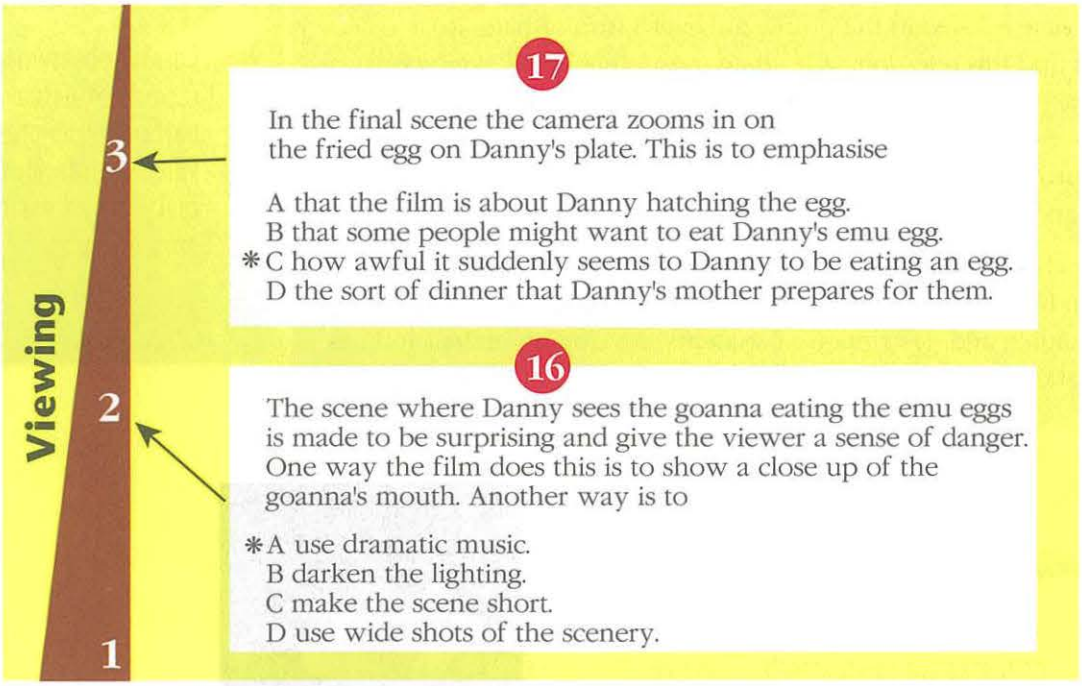
Students' performances on Question 16 show that they find this question about the dramatic use of music relatively easy. Although this outcome is listed in the profile at Level 3 (top of page 16), students find this question easier than most other Level 3 questions and as easy as most Level 2 questions. This observation suggests that students can identify the symbolic use of music in films at profile Level 2 provided that they simply have to recognise its use in a particular instance.

By analysing students' responses to other questions about music and sound effects in films it may be possible to construct a still better understanding and description of students' developing understandings of dramatic devices.

Careful observations and records of actual student performances provide valuable information for revising and enriching a progress map.



The Viewing strand of the English profile for Australian schools locates the ability to identify the symbolic use of music in films at Level 3.



Student performances locate Question 16 at Level 2 on the Viewing strand suggesting that students can identify the symbolic use of music in films at profile Level 2 when they simply have to recognise its use in a particular instance.

Progress Maps

Path of typical progress

A progress map (or developmental continuum) describes the path of typical student progress through an area of learning. It is a shared framework for monitoring student growth.

Attainment levels

It is common to visualise student progress in an area of learning as occurring along a continuum from lower-level, rudimentary skills and understandings to higher-level, more sophisticated skills and understandings. Along such a continuum it is possible to specify a number of broad and fuzzily-defined levels of attainment.

Constructing a progress map

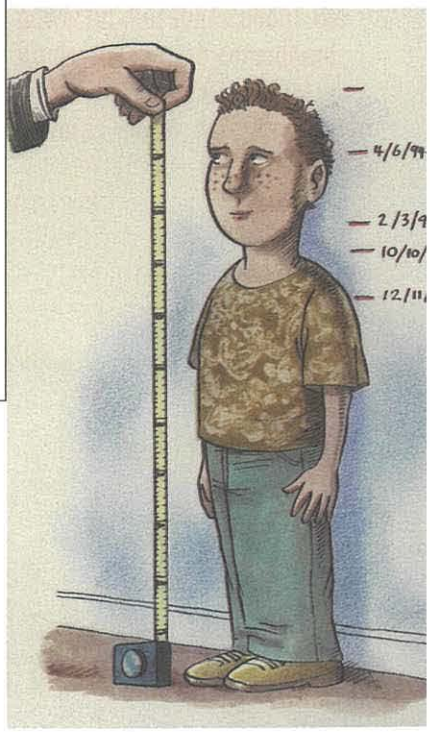
Some progress maps are constructed by asking teachers and other subject matter specialists to draw on their knowledge of typical paths of student progress through a learning area; other maps are constructed from an analysis of students' performances on assigned tasks. The ideal may be an approach which combines expert judgement and empirical data.

Verifying a progress map

When the first draft of a progress map is based on expert opinion, the resulting draft must be tested against practice. Is the order of outcomes on the map consistent with typical classroom observations? This process is similar to taking a road map drawn from memory and checking its accuracy against the terrain it is intended to represent.

Enriching a progress map

A progress map must be constantly checked, updated and enriched. Careful observations and records of actual student performances provide valuable information for revising and enriching a progress map.



1 Western Australian Ministry of Education (1991) *First Steps Spelling Developmental Continuum*, Perth W.A.: Ministry of Education.

2 Curriculum Corporation (1994) *Introducing Statements and Profiles*, Carlton: Curriculum Corporation, p. 8.

3 Campbell Hill, B. & Ruptic, C. (1994) *Practical Aspects of Authentic Assessment: Putting the Pieces Together*, Norwood, MA.: Christopher-Gordon Publishers Inc, p. 242.

4 Western Australian Ministry of Education.

5 Larter, S. (1991) 'Benchmarks: Toronto's Response to the Testing Problem', *Newsletter, Federation of Women Teachers' Associations of Ontario*, Vol. 10, No. 1, p. 5.

6 Glaser, R. (1963) 'Instructional Technology and the Measurement of Learning Outcomes: Some Questions', *American Psychologist*, Vol 18, pp. 519-521.

7 Western Australian Ministry of Education, p. iv.

8 Campbell Hill & Ruptic, p. 235.

9 Black, P. (1987) *Report of Task Group on Assessment and Testing* London: Department of Education and Science, p. 2.

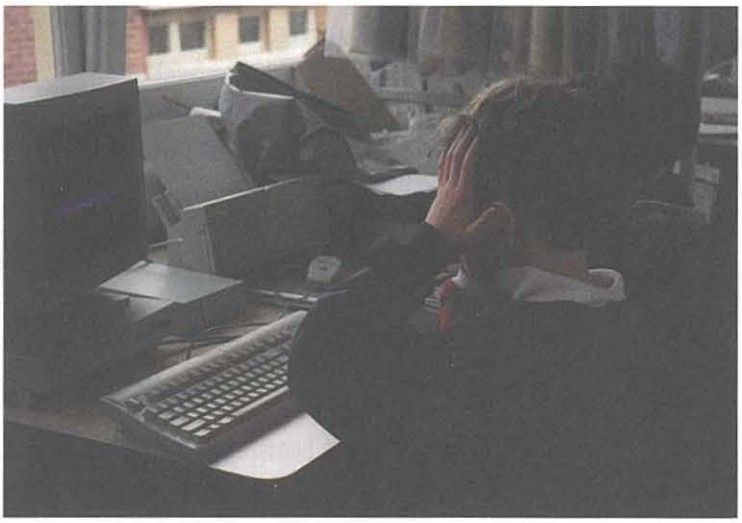
10 Curriculum Corporation, p. 10.

11 Curriculum Corporation (1994) *Mathematics Profile for Australian Schools*, Carlton: Curriculum Corporation, p. 58.

12 Whetton, C. (1992) 'The Assessment System: Purposes and Constraints', Paper delivered at AERA, San Francisco, 1992, p. 2.

13 Larter, p. 5.

14 Forster, M., Mendelovits, J. & Masters, G.N. (1994) *DART Developmental Assessment Resource for Teachers*, Melbourne: Australian Council for Educational Research.



assessment methods

The evidence (observations) used to estimate students' levels of attainment on a progress map can be based on a variety of *assessment methods*. These assessment methods include portfolios, performances, projects, products, and paper and pen assessments.

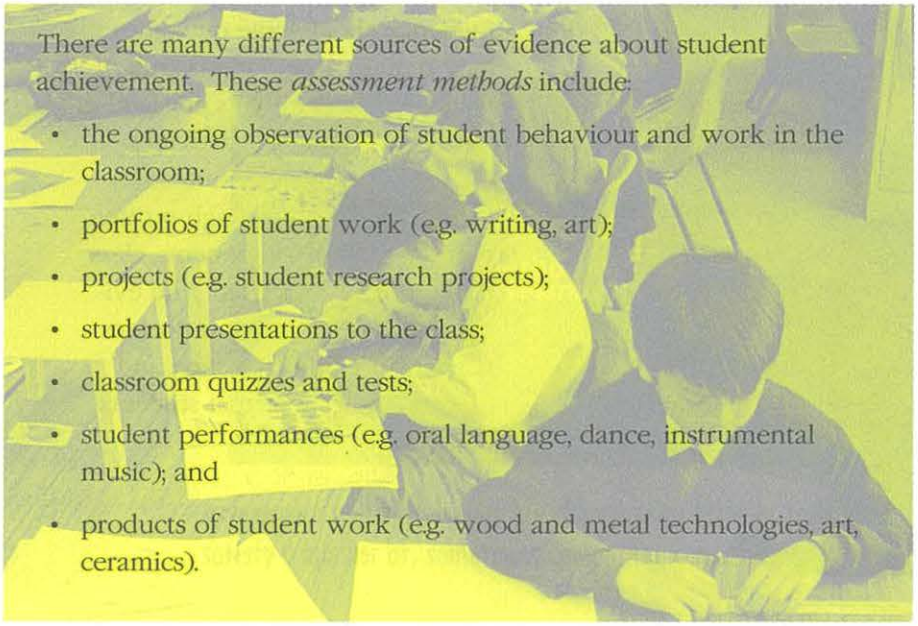
In planning the collection of evidence it is important that assessment methods are chosen carefully. Not all methods are capable of providing information about all learning outcomes. Some outcomes require particular methods. The assessment of a student's abilities to find and select relevant information, to analyse collected material, and to write a report of their conclusions is probably best done through an assigned project, for example. The assessment of a student's ability to write for a range of purposes and audiences is probably best done through a collection (portfolio) of their writing. And the assessment of a student's abilities in instrumental music is probably best done by observing musical performances.

A complete picture of student achievement in an area of learning depends on useful information about a broad range of learning outcomes. An assessment program which addresses only some outcomes provides a limited picture of achievement and so provides incomplete feedback to teaching and learning.

A second reason for choosing assessment methods carefully is that the choice of one method over another can send powerful messages to students about what is valued. If practical laboratory skills are considered an important aspect of school science, for example, but course results are based entirely on paper and pencil tests, then students—and possibly teachers, too—will focus their efforts only on outcomes that can be assessed through paper and pencil tests. If the methods used to collect evidence address only some outcomes, then assessment procedures can distort teaching and learning.

In most areas of learning, no single assessment method is capable of providing evidence about the full range of learning outcomes.

Assessment methods send powerful messages about the kinds of learning considered worthy of recognition and reward.



There are many different sources of evidence about student achievement. These *assessment methods* include:

- the ongoing observation of student behaviour and work in the classroom;
- portfolios of student work (e.g. writing, art);
- projects (e.g. student research projects);
- student presentations to the class;
- classroom quizzes and tests;
- student performances (e.g. oral language, dance, instrumental music); and
- products of student work (e.g. wood and metal technologies, art, ceramics).

deciding on assessment methods

In an ideal world, the methods used to collect evidence of achievement in an area of learning would be carefully designed to:

- reflect curriculum priorities;
- provide feedback that informs and guides instruction;
- be fair to all students;
- provide results that are reliable and comparable across students; and
- be administratively convenient and inexpensive.

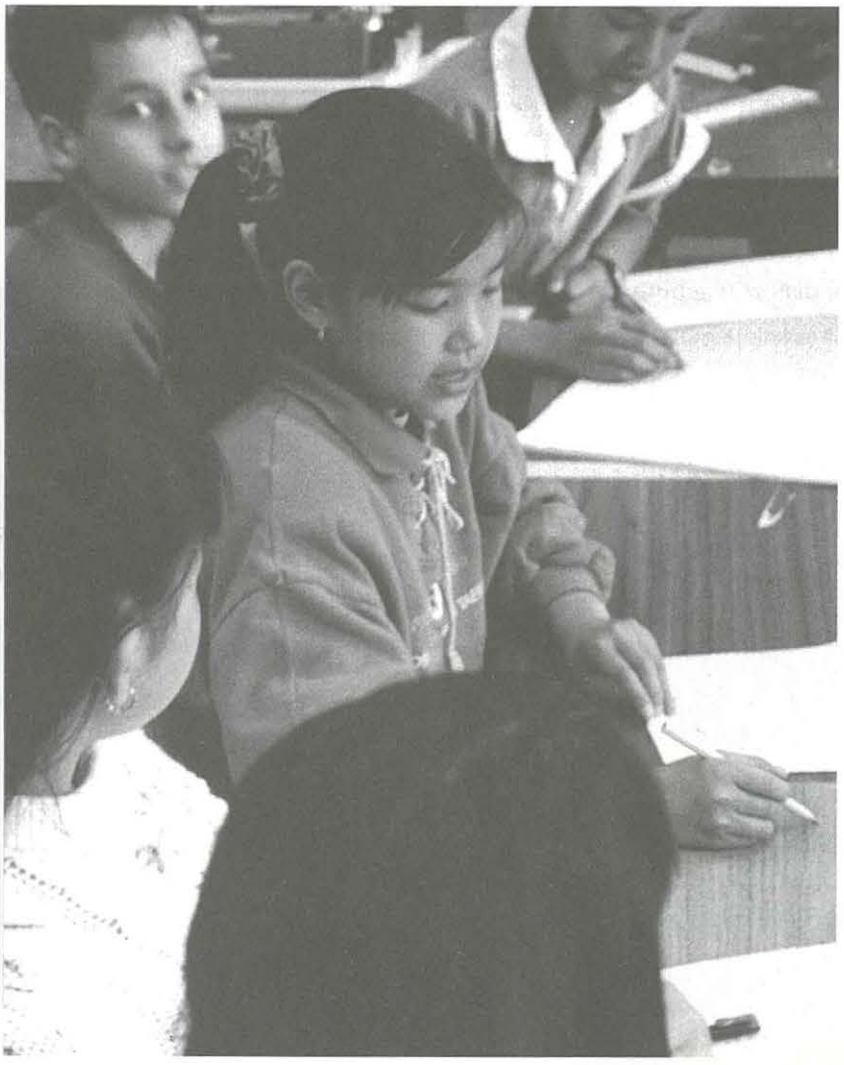
Assessment methods should be chosen to faithfully reflect and provide evidence about the range of knowledge, skills and understandings that make up an area of learning.

In practice, methods of assessment usually are decided on pragmatic grounds. For classroom purposes, useful feedback to teaching and learning, practical convenience, and the availability of time and materials may be judged more important than considerations of reliability, fairness and comparability. In an admissions program, on the other hand, objectivity and fairness may be judged more important than the accurate reflection of all curriculum priorities or the provision of instructionally useful feedback.

curriculum relevance

In most educational assessment programs, the matching of assessment methods to curriculum goals is fundamental. Assessment methods are chosen to faithfully reflect, and provide evidence about, the range of knowledge, skills and understandings that make up an area of learning. Conclusions about student achievement are *valid* only when based on evidence about this full range of outcomes.

The collection of evidence in an area of learning usually requires a variety of assessment methods. Evidence about some outcomes may require observations and judgements of students engaged in day-to-day classroom activities. Evidence about other outcomes may be collected through written work including assignments, classroom quizzes and tests. And evidence about still other outcomes may require the inspection and judgement of items that students have made.



In recent years many large-scale assessment programs have been reconstructed to provide evidence about a broader range of curriculum goals than have been addressed in the past through written tests. These programs sometimes include portfolios of student work and assessments of oral communication skills, research skills, laboratory skills, and practical and manual skills.

In assessment contexts in which high levels of reliability and comparability are important, curriculum coverage may be considered less important than ensuring that all students are assessed on a 'level playing field' using common tasks and identical assessment criteria. When assessments are made for the purposes of selection, certification, or the award of scholarships, the requirements of fairness and objectivity may limit the range of assessment methods and hence the learning outcomes addressed. In the past, extensive use has been made of paper and pencil tests and examinations in these assessment contexts because of their perceived objectivity and reliability.

Practical convenience and cost also sometimes take precedence over curriculum fidelity. In large assessment programs in which tens or hundreds of thousands of students are assessed at a time, it may not be practicable to use some methods of assessment, meaning that evidence may be collected about some learning outcomes but not about others.

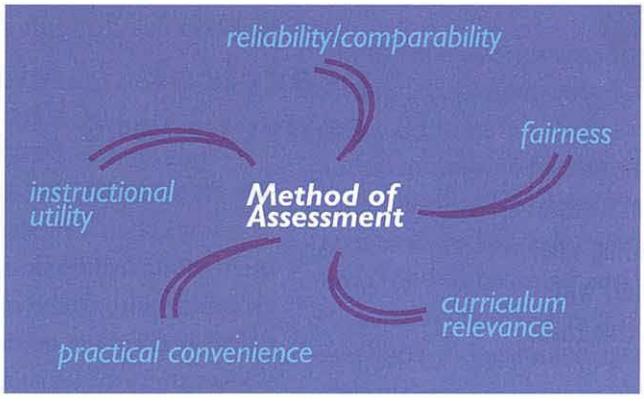
instructional utility

Ideally, methods for collecting evidence of student achievement should be designed to provide information which is instructionally useful.

Assessment methods should be chosen and designed to yield insights into students' knowledge and conceptual understandings which can be used in future teaching and learning. In practice, feedback to instruction is less important in some assessment contexts than in others.

One form of feedback likely to be useful in planning teaching and learning is a general estimate of where an individual is in their development: What level of attainment have they reached? What kinds of skills, understandings and knowledge are typical of students

The most instructionally-useful methods of assessment may be methods that provide information about how individuals are thinking.



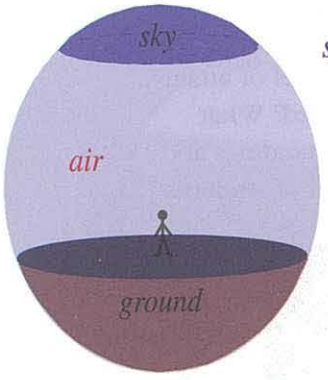
Assessment methods usually are chosen in an attempt to satisfy a number of, sometimes competing, considerations.

at that level of attainment? What learning activities are likely to be most appropriate and useful next?

Another form of feedback is a more detailed analysis of a student's performances including an identification of special strengths and weaknesses.

But perhaps the most instructionally useful methods of assessment are those that provide information about how individuals are thinking. What understandings have students constructed for themselves? What pictures do they carry in their heads? How do these understandings differ from the understandings that teachers are trying to develop?

Assessment methods designed to collect information about how students are thinking can be quite different from methods designed to establish whether or not students have memorised and can recall specific facts, can apply taught procedures, or demonstrate particular skills.



When asked to draw the Earth, primary school children draw different kinds of pictures, in this case integrating what they 'know' (the Earth is flat) with what they have been taught (the Earth is round).[1]

The collection of evidence about how students are thinking requires opportunities for students to explain in their own words and drawings their understandings of the material they are learning. The purpose of assessment methods of this kind is to provide insights into students' conceptual understandings.

fairness

Assessment methods must be fair to all students. This requirement is also more important in some assessment contexts than in others. It is especially important that assessments are made without prejudice in high stakes contexts where results can influence the allocation of resources or students' chances of employment or admission to educational programs. But fairness is an important consideration in all assessment contexts.

Teachers and other assessors need to be aware of the ways in which the choice of an assessment method can influence the performances of students who do not speak English as a first language, who come from particular cultural backgrounds, or who are physically disadvantaged. They also need to be aware of the ways in which the choice of an assessment method can influence the relative performances of girls and boys.

practical convenience

Another desirable feature of an assessment method is that it is convenient to implement. Practical convenience is especially important when there is a desire to collect evidence on a wide variety of learning outcomes for all students. When this is the intention, it is important that assessment procedures are manageable, easily incorporated into usual

classroom activities, and capable of providing information that justifies the time and money required.

Practical feasibility takes on an even greater significance in assessment programs which attempt to provide both comprehensive curriculum coverage and high levels of reliability and comparability. In practice, these three intentions often prove difficult to satisfy simultaneously.

In Britain, for example, an attempt was made to collect evidence across the curriculum, for all students at particular Year levels, and in a way that would allow reliable comparisons across teachers and schools. Complex 'standard assessment tasks' matched to curriculum goals were developed for administration by teachers. But these tasks took, on average, many hours of class time to administer, and tens of millions of dollars to develop. And, despite this effort, questions remained about the comparability of the resulting assessments because of differences in the conditions under which tasks were completed and the subjectivity of teachers' judgements.[3]

> The [national] standard assessment tasks were constructed to be as educational as possible. In line with today's aims for primary schooling, the process is seen as being just as important as the product. So the tasks were designed to be not too far away from what children are doing anyway. 'Teaching to the test' is avoided, children are not unnerved by too big a change in routine, and teachers can use their skills of assessment and diagnosis.
> These laudable aims have made the standard assessment tasks time-consuming and complicated.
>
> *Times Educational Supplement*[2]

reliability / comparability

Another word for reliability is 'accuracy'. How accurately does the collected evidence reflect an individual's skills, knowledge and understandings?

Reliability is closely linked to considerations of fairness. To the extent that an assessment is not an accurate reflection of a student's achievements, it can be unfair to the student concerned.

Reliability is particularly important when assessments are used to make comparisons of individuals or groups. If the assessment of a student's level of attainment in an area of learning does not provide an accurate picture of his or her achievement, then any decision based on that assessment is likely to be unfair. Decisions about graduation, course admissions, and the award of scholarships, for example, depend on reliable information. So do decisions to report students' achievements on public certificates.

Reliable evidence also is important in assessment programs designed to monitor standards across an education system over time. Because changes in average levels of student achievement over time tend to be small and difficult to detect, sensitive assessment methods and accurate data are essential.

The desire to place all students on an equal footing and to assess their performances in an objective way

If the assessment of a student's level of attainment in an area of learning does not provide an accurate picture of his or her achievements then any decision based on that assessment if likely to be unfair.

against common criteria has resulted in the extensive use of paper and pencil tests and examinations. Objectively-scored paper and pencil tests became particularly popular during the 20th Century because of the perceived reliability and comparability of test scores.

But in recent years efforts have been made to develop procedures to improve the reliability of evidence about learning outcomes that cannot be addressed through objectively-scored paper and pencil tests. In the use of methods such as portfolios, performances, projects, and products of student work as sources of evidence, reliability can be enhanced by

- specifying the kinds of evidence to be collected;
- specifying the criteria to be used in assessing student work;
- training markers in the use of criteria;
- providing samples of students' work illustrating the assessment criteria;
- identifying and adjusting for differences in the standards applied by different assessors; and
- identifying and adjusting for differences in the details of the tasks that students attempt.

Examinations were introduced as a response to concerns about nepotism and prejudice in the selection of Chinese civil servants more than 1500 years ago.

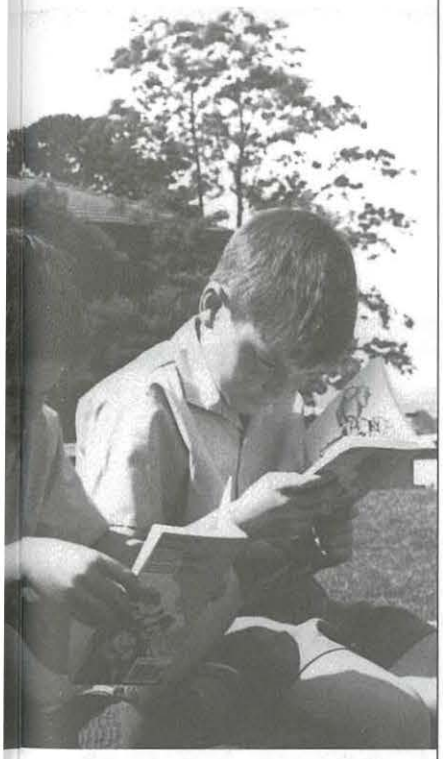
For Discussion...

'Because the objective test item provides a more uniformly standardized, carefully controlled process of measurement than does the essay test, it is, and should be, regarded as a technical advance in the measurement of educational achievement.'

Robert Ebel[4]

Patricia Broadfoot argues for abandoning 'carefully designed, objective' assessment techniques in favor of approaches that will 'promote the right kinds of educational processes, however suspect psychometrically such approaches may be.'

Patricia Broadfoot[5]



Some Assessment Methods

Portfolios

Portfolios are collections of student work assembled over a period of time. These may be collections of day-to-day work (Working Portfolios); collections of work for assessment but documenting the processes used to develop items in the portfolio (Documentary Portfolios); or selections of students' best work for summative assessment (Show Portfolios).

Performances

Student performances must be observed and assessed as they occur or audiotaped or videotaped for assessment. Examples include oral presentations, dance, gymnastics, and instrumental music.

Projects

Projects are undertaken over a period of time and often involve the collection and analysis of data and the preparation of a written report. Posters are sometimes used by students to report the findings and conclusions of their investigations.

Products

Products are items made by students. Examples include pieces of artwork (drawings, paintings, sculptures), items of food, articles made of wood, metal, plastic, and ceramics. In each case the product of a student's work is available at the completion of the process for assessment.

Paper and pen assessments

Paper and pencil tests usually are completed in a limited period of time under specified conditions. Questions on paper and pencil tests can take many forms, but the most common are short answer, essays, and multiple-choice questions.

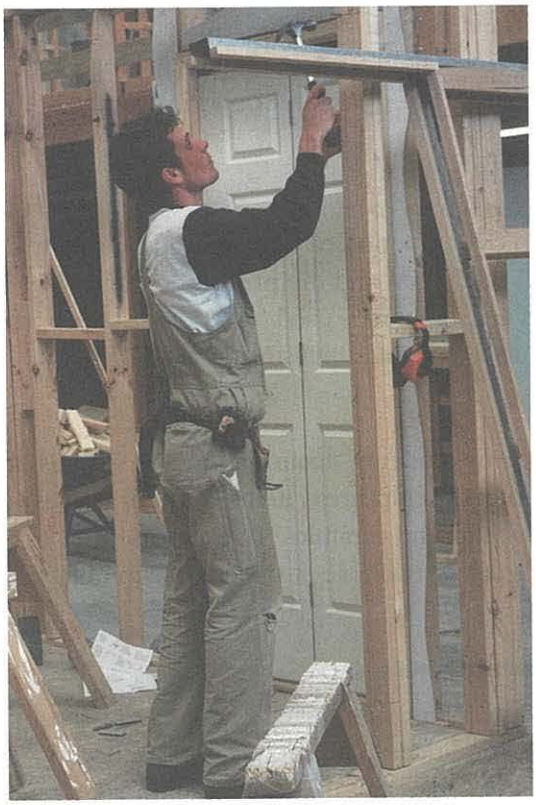
1 Nussbaum, J. (1979) 'Children's Conceptions of the Earth as a Cosmic Body: A Cross-Age Study', *Science Education*, Vol. 63, No. 8, pp. 3-93.

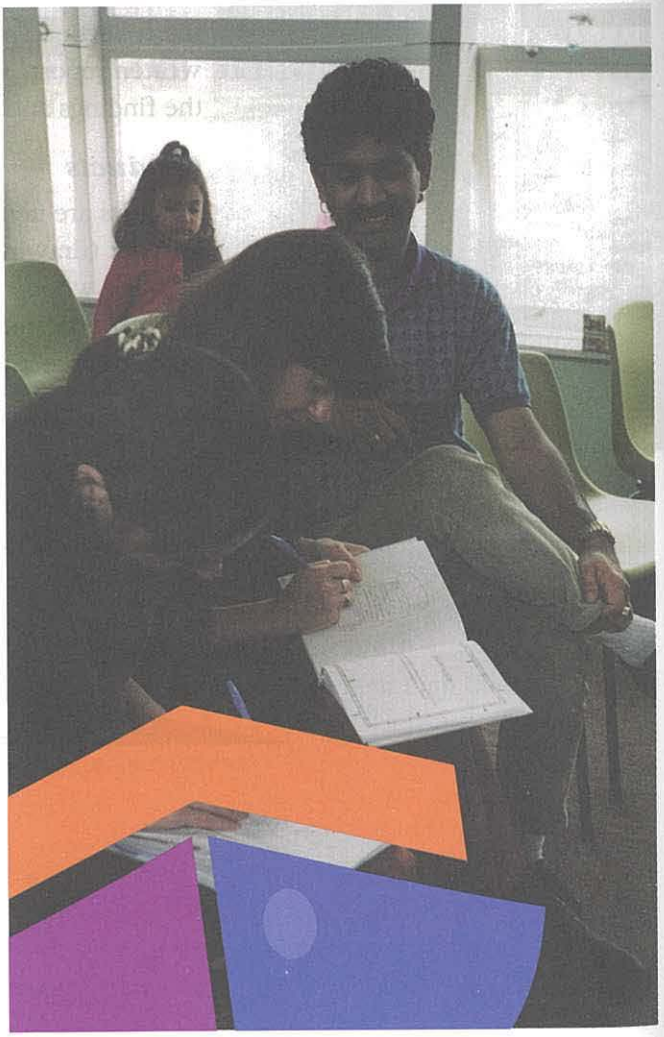
2 *Times Education Supplement*, 5 May, 1991, p. 13. © Times Supplements Limited, 1991.

3 Patricia Broadfoot and Marilyn Osborn questioned the reliability of SAT assessments as part of their Primary Assessment Curriculum and Experience (PACE) project.

4 Ebel, R. (1965) *Essentials of Educational Measurement*, New Jersey: Prentice-Hall, p. 132.

5 Broadfoot, P. (1992) 'Toward Profiles of Achievement: Developments in Europe', in Eckstein, M. A. & Noah, H. H. (eds) *Examinations: Comparative and International Studies*, Oxford: Pergamon, p. 69.





judging and recording

An important feature of developmental assessment is the making and keeping of *written records* of student behaviour as a basis for estimating levels of attainment on a progress map.

Some of the records teachers keep are based on observations that they happen to make in the course of their teaching.

Other records are based on planned and purposeful observations of students' performances on assigned activities and work. In these more formal assessment settings, teachers use assessment methods such as portfolios, performances, projects, products, and paper and pen tasks as sources of evidence. Students' responses and work usually are then judged in some way (e.g. rated, categorised, scored).

Records of observations and judgements are the evidence needed to estimate students' levels of attainment in an area of learning. This evidence is always limited, either to the observations that teachers happen to make, or by the limited number of activities and tasks on which performances are observed.

anecdotal records

Teachers have numerous opportunities to make observations of student behaviour in their day-to-day teaching. These observations provide a basis for making judgements about the stages that individuals have reached in their learning, for identifying special strengths and weaknesses, and for evaluating the progress that students have made over time.

But these observations tend to be casual and fortuitous rather than systematic and planned. Day-to-day observations are easily forgotten if not recorded. And in a busy classroom, quiet students sometimes go unnoticed, meaning that their achievements are less likely to be appreciated and their difficulties less likely to be detected.

Lower achieving students, in an effort to avoid embarrassment, often develop strategies to disguise their difficulties. Some students feign the ability to read, for example, and are successful in disguising their reading difficulties year after year.

Beyond this, casual and fortuitous observations may provide evidence about student achievement in only some areas of the curriculum. Day-to-day observations of a non-strategic kind are unlikely to provide evidence about individuals' progress across the full range of outcomes in an area of learning.

The usefulness of day-to-day classroom observations as a basis for estimating and monitoring levels of attainment in an area of learning can be enhanced by:

- ensuring that relevant observations are made for *all* students in a class;
- ensuring that observations are made in relation to the *range* of important learning outcomes; and
- keeping a *written record* of these observations.

Teachers use a variety of schemes for recording observations and judgements, ranging form anecdotal records to procedures for rating and scoring students' work.

Lisa Norwick in Seattle uses an observational system based on the use of post-it notes. She writes each student's name on a blank post-it note and sticks these to a clipboard. As she moves around the classroom observing student activities and talking with students about their work she records her observations of each student on that student's post-it note.
Lisa's observations are focused on outcomes from the developmental continua in reading and writing that she uses as frameworks for monitoring student progress in these areas of learning. She later transfers each student's post-it note to a record book, placing it alongside other post-it note observations for that student.

Various schemes can be used to make classroom observations more focused and systematic.

A variety of schemes for making and keeping anecdotal records have been developed by teachers. Some of these use card systems, others use computer software. But the best of these schemes encourage the collection of information about each student in a class, on a range of learning outcomes, and in the form of a written record.

There are two general ways of rating student work: analytically and holistically.

rating scales

Judgements of student work also can be made using rating scales. In using a rating scale teachers judge the quality of a piece of student work against specified criteria.

When work is rated *analytically,* teachers assess student work on each of a number of specified criteria. To make these judgements, teachers must consider the piece of work from a number of different perspectives separately. A set of criteria for judging a Studio Arts folio (see page 29), for example, might require separate assessments on such criteria as planning and preparation, use of materials, expression of ideas, and use of imagination.

Various *rating categories* can be used to record judgements of the quality of student work. In the Studio Arts example, folios are rated High, Medium, Low, or Not Shown on each of the ten criteria. Other categories sometimes used in analytic ratings of student work include Almost Always, Sometimes, and Hardly Ever.

Analytic rating schemes are most useful when accompanied by annotated examples of their use. Greater consistency of ratings is likely to be achieved when teachers have access to examples of student work to illustrate the categories High, Medium, and Low on each criterion.

Student work also can be rated *holistically.* In holistic ratings, a single set of rating categories is developed and used to make an overall judgement of the quality of a piece of work.

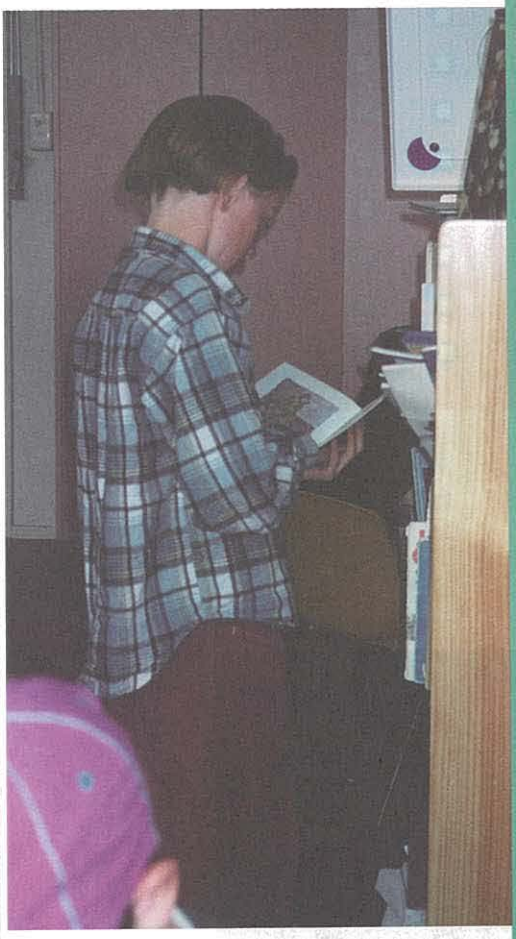
An example of a set of holistic rating categories is the rating scale developed as part of the Riverside Curriculum Assessment System for judging Year 8 students' responses to the poem 'Daffodils' by William Wordsworth (see page 30). These rating categories were developed to assist teachers to focus on students' understandings of and

CRITERIA	High	Med	Low	Not Shown
1 suitable definition and explanation of the studio works to be undertaken	☐	☐	☐	☐
2 suitable planning and preparation to implement the studio works	☐	☐	☐	☐
3 sensitive use of media and materials	☐	☐	☐	☐
4 skilful use of equipment and tools	☐	☐	☐	☐
5 competent application of studio techniques	☐	☐	☐	☐
6 effective communication and/or expression of ideas	☐	☐	☐	☐
7 a personal and imaginative approach appropriate to the ideas of the studio works	☐	☐	☐	☐
8 aesthetic qualities of the studio works	☐	☐	☐	☐
9 finish suitable to the studio works	☐	☐	☐	☐
10 implementation of the ideas outlined in the work brief/s	☐	☐	☐	☐

Criteria used for the analytic rating of a Studio Arts Folio.[1]

approaches to this poetry task. The holistic rating scale describes increasing levels of performance on a limited number of criteria: response to the function of imagery in the poem, organisation and focus of students' written responses, and grammar, punctuation and mechanics.

Holistic rating scales are most useful when accompanied by annotated examples of student work at each level on the described scale.



Poetry

Exercise

The poet uses a great deal of imagery (mental pictures) in the poem. For example, he states that there are a great number of golden daffodils fluttering and dancing, and he compares them to the stars in the Milky Way. How do these mental pictures add to your understanding or enjoyment of the poem? Write a paragraph explaining your answer.

Scoring Rubric

Note: At Levels 3 and 4, errors in mechanics and grammar (e.g. fragments, misspellings, flawed punctuation, and incorrect capitalization) should not impede understanding.

Assign points to student responses that most closely match the characteristics listed.

4
- provides a direct, accurate response to the function of imagery
- shows good organization and a clear focus
- includes few, if any, errors in grammar, usage, or mechanics

3
- exhibits a fairly clear and logical response to the question
- contains minor organizational flaws or a somewhat unclear focus
- may include some errors in grammar, usage, or mechanics

2
- attempts to address the question
- includes confusing organization and a focus which is unclear
- contains problems in mechanics that interfere with communication

1
- barely attempts to address the question
- includes little organization and is not focused
- complicates message with serious problems in language and mechanics

0
- indicates that the student has failed to attempt the question

N/S (nonscorable)
- indicates that the response is illegible or unreadable

Riverside Curriculum Assessment System

Scale for making holistic ratings of students' written responses to a poetry task.

partial credit

Partial credit scoring is a method for recording the steps students successfully complete in the solution of a problem or their partially correct understandings and strategies.

Many students who are unable to successfully complete assigned tasks and problems display some levels of understanding and success. Rather than recording all unsuccessful attempts at a problem as 'wrong', partial credit scoring recognises and records students' varying levels of partial success.

In mathematics, for example, it is well known that students do not simply make mistakes at random, but often fail to solve problems because they have partially correct but incomplete understandings of mathematical concepts and processes. The superficial errors that students make often are evidence of incomplete understandings of concepts.

In her research into students' understandings of fractions, Kikumi Tatsuoka identified a number of errors that students make when asked to solve fractions addition problems such as

$$2\frac{1}{8} + 3\frac{5}{6} = ?$$

The errors students make often reflect different levels of understanding (see page 32).[3] Some students make errors that suggest little if any understanding of fractions and their addition. They frequently resort to 'rules' that appear to have little meaning for them. Other students who are unable to provide the correct answer nevertheless complete some steps correctly and arrive at answers which suggest some understanding.

Rather than treating all incorrect responses to a task such as this in the same way (i.e. 'wrong'), it is sometimes useful to distinguish among the different responses that students give by categorising them according to degree of completeness or apparent level of understanding.

Partial credit scoring recognises and records students' varying levels of success in completing tasks and solving problems.

dichotomous records

Another approach is to record whether or not a student successfully completes a task. Records of this kind are called 'dichotomous' because there are only *two* categories for recording observations.

One context in which dichotomous records of student performance are kept is in systems of 'work required' assessment where teachers and students negotiate the work to be done and sign off on each piece of work as it is completed. In systems of 'work required' assessment, the work to be completed is specified in terms of the processes that students must satisfactorily complete.

A disadvantage of 'work required' systems is that by focusing on a set of activities to be satisfactorily completed, they may not explicitly address or provide useful information about student progress in relation to learning outcomes.

A second context in which it is common to record students' performances dichotomously is in paper and pencil tests in which a judgement is made about the correctness or acceptability of students' answers to each question.

completed	not completed
acceptable	not acceptable
right	wrong

In dichotomous scoring teachers judge whether students' performances meet the criterion for completeness, acceptability or correctness.

Adding Fractions

4 Adds simple fractions correctly. Demonstrates an awareness that only like fractions (same denominator) can be added. Converts to like fractions correctly. Obtains correct answer.

e.g. $2\frac{1}{8} + 3\frac{5}{6} = 5 + \frac{3}{24} + \frac{20}{24} = 5\frac{23}{24}$

3 Adds simple fractions correctly. Demonstrates an awareness that only like fractions (same denominator) can be added. Converts to like fractions correctly. Makes error at final step.

e.g. $2\frac{1}{8} + 3\frac{5}{6} = 6 + \left(\frac{3}{24} + \frac{20}{24}\right) = 6\frac{23}{24}$

2 Adds simple fractions correctly. Demonstrates an awareness that only like fractions (same denominator) can be added. Does not demonstrate an understanding of how to convert to like fractions and makes serious errors of understanding in the attempt.

e.g. $2\frac{1}{8} + 3\frac{5}{6} = 2\frac{1}{24} + 3\frac{5}{24} = 5\frac{6}{24}$

1 Does not add simple fractions correctly. Does not demonstrate an awareness that only like fractions (same denominator) can be added. May convert mixed to improper fractions correctly, but shows no understanding of the reason for doing this. Shows some elementary understanding of fractions (e.g. by attempting to sum integers and fractions separately). May use rules like those for multiplying fractions.

e.g. $2\frac{1}{8} + 3\frac{5}{6} = \frac{17}{8} + \frac{23}{6} = \frac{17}{8} + \frac{6}{23} = \frac{17 \times 6}{8 \times 23}$

0 Does not add simple fractions correctly. Response reflects little understanding of fractions (e.g. treats integer, numerator, denominator equivalently).

e.g. $2\frac{1}{8} + 3\frac{5}{6} = \frac{2+1+8}{3+5+6} = \frac{11}{14}$

Kikumi Tatsuoka has identified different kinds of errors that students make when solving fraction addition problems. The errors shown here reflect different levels of understanding.[4]

In the TORCH *Tests of Reading Comprehension*, for example, students' written answers to each question are judged as either 'acceptable' or 'unacceptable' using a provided key[5]. Students first read a passage of text and then complete a 're-telling' of the passage by filling in strategically-located gaps (see page 34). The re-telling is in somewhat simpler language than the original passage, and the gaps are located to assess students' understandings of important ideas in the passage.

Each TORCH key provides examples of acceptable and unacceptable responses based on answers given by several hundred students during the construction of the tests. Teachers using TORCH read students' written responses and make a judgement about the acceptability of each answer using the provided key.

judgements of outcomes 'achieved'

A fifth method involves a decision about whether or not individuals have 'achieved' particular learning outcomes.

In using this method, teachers make a list of relevant outcomes and then make a judgement about whether or not a student has achieved each outcome on the list.

The difficulty in making judgements of this kind is in knowing how much evidence is required, and in what range of contexts, before deciding that a student has 'achieved' an outcome.

Leon Deleuil and Cliff Malcolm point out that, in general, a conclusion about the achievement of an outcome is not possible on the basis of a single piece of work or a single activity (apart from a comprehensive test) but requires an on-balance judgement based on observations made in a range of contexts.[6] They suggest that judgements about whether or not individuals have 'achieved' an outcome involve four considerations:

1. Has the student covered the content implicit in the outcome?
2. Can the student apply the knowledge and skills accurately in different contexts?
3. Were the tasks fair and challenging in terms of the student's background, language or any special circumstances?
4. Is the standard of the student's work consistent with the profile level to which the outcome belongs?

Teachers who use this method to record their observations and judgements of student achievement sometimes mark the outcomes that individuals have achieved using a highlighter pen. These highlighted outcomes then provide the basis for making on-balance judgements about students' locations on the progress map that the outcomes define.

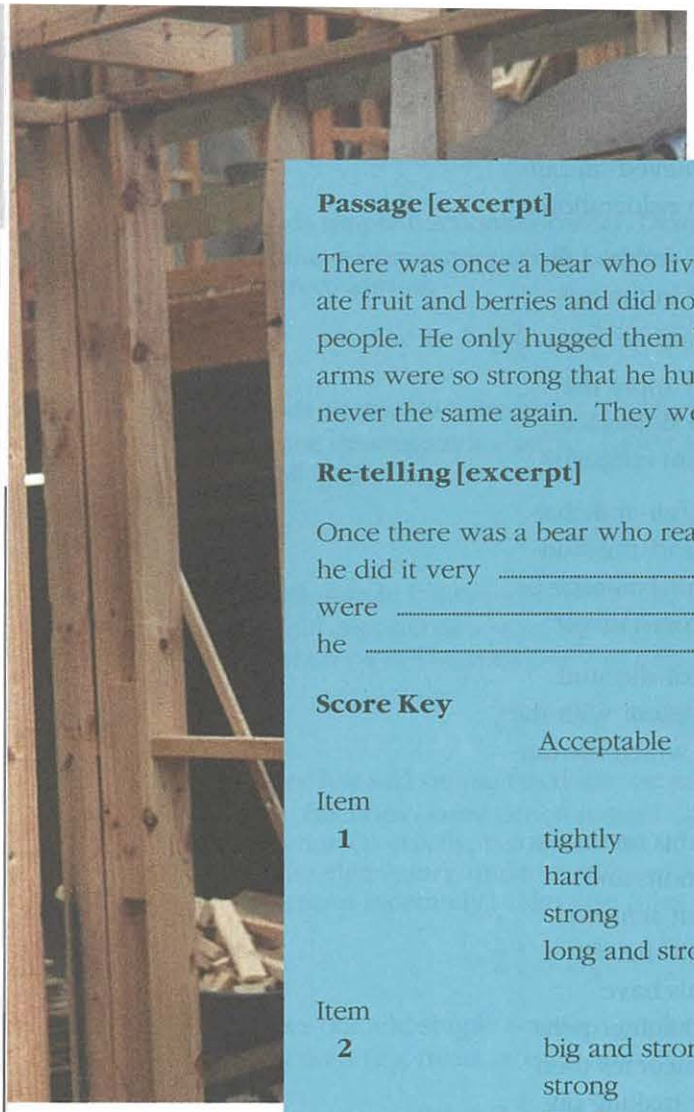
If Janine produces a poster on tin mining and talks about her work to the class clearly and confidently, does that mean she has achieved the level 2 outcome 'Describes changes that occur in the local environment'?

How much more information do we need?

Is this event typical of her performance?

Are other tasks and other contexts required for the outcome?

Using the Science Profile[7]



Passage [excerpt]

There was once a bear who lived in a cave in the mountains. He was a mountain bear. He ate fruit and berries and did no harm to anybody, but he had one bad habit. He would hug people. He only hugged them because he liked them, but they did not know that. His furry arms were so strong that he hugged much too tightly. Some of the people he hugged were never the same again. They were quite flat when he let them go, and lop-sided.

Re-telling [excerpt]

Once there was a bear who really liked hugging people. When he hugged them he did it very ..... because his furry arms were ..... . He hugged them because he **1** **2** **3**

Score Key

	Acceptable	Unacceptable
Item **1**	tightly hard strong long and strong	well much bad
Item **2**	big and strong strong tough big strong and furry	so strong that he killed them hard long outstretched tight furry
Item **3**	liked them loved people liked hugging people liked hugging loved it had a habit of doing it had a habit	loved to be hugged was kind loved

The TORCH Tests of Reading Comprehension require teachers to read and judge the acceptability of students' answers to each question. The key provides examples of acceptable and unacceptable answers.

categorising student responses

A final method for recording observations and judgements of student performance is to allocate individuals' responses to one of several pre-defined categories of response to a task.

A number of different techniques can be used to construct a set of response categories for a task. One technique—known as 'phenomenography'—has been developed by Ference Marton in Sweden.[8] Marton and his colleagues present students with an open-ended task, question, or problem designed to provide information about their understandings of an aspect of a learning area. These tasks, questions or problems usually are presented in interviews during which students are encouraged to 'think out loud'.

Students' explanations are tape-recorded for later detailed analysis. Work of this kind has shown that, whatever the area of learning, students' explanations usually display a small number of different ways of thinking about a problem. These qualitatively different understandings are organised into a set of 'categories' representing increasingly sophisticated levels of understanding.

> Phenomenography is a research method for mapping the qualitatively different ways in which people experience, conceptualize, perceive, and understand various aspects of, and phenomena in, the world around them.
>
> *Ference Marton*[9]

Marton provides an example of a set of categories constructed in this way. In an earlier study, two researchers Andersson and Kärrqvist had shown that very few Year 9 students in Sweden understood basic properties of light and the relationship between light and seeing: 'Teachers probably do not systematically teach this fundamental understanding, which is so much a part of a teacher's way of thinking that they neither think about how fundamental it is, nor recognise that it can be problematic for students'.[10]

In his study, Marton gave students a question about light and seeing (see page 37) and analysed their verbal and written explanations. He found five different kinds of understandings in students' responses. These varied from (1) no apparent understanding of the role of light in seeing to (5) an understanding that light must be reflected from an object to the eyes in order for it to be seen.

Categories of understanding constructed in this way can be used by teachers to categorise students' responses to assigned tasks. When mapped on to a progress map (e.g. 'conception (2) is typical of students at Level 1') response categories of this kind can be an aid to estimating students' levels of attainment on that map.

Students' explanations usually display a small number of different ways of thinking about a problem.

common 'errors'

In making and recording observations and judgements of student achievement, teachers need to be aware of biases and 'errors' commonly found in judgements of human performances. These 'errors' are well documented in the assessment literature. They include:

Pre-judging When teachers are able to observe a student over a period of time, they usually develop expectations of that student's performances. These expectations can become a source of error in assessment if they lead to judgements based on perceptions of student ability rather than actual student performances.

Raters who develop systematic expectations regarding the performance of a ratee may find it difficult to accurately evaluate that person's performance if he or she departs from a previous pattern of performance.

Murphy et al.[11]

Confusing achievement with effort Teachers sometimes use assessments to reward effort. A desire to reward a special effort can result in the award of a higher level of achievement than a student's performance warrants. Assessments are also sometimes used to 'jolt' students believed to be working below capacity. The use of assessments to send 'messages' to students may undermine their validity and reliability as measures of achievement.

Different standards for different students Prior impressions of student abilities can lead teachers to apply different standards when assessing student performances.

Whenever teachers modify their standards on the basis of prior impressions or beliefs about students, there is a risk that their assessments will not accurately reflect individual abilities.

Cultural stereotyping A number of researchers have reported systematic tendencies for assessors to assign higher performance assessments to individuals of their own race.[12] In an analysis of 74 studies of race effects on performance evaluations, Kraiger and Ford reported a consistent tendency for white assessors to rate white ratees higher than black ratees, and black raters to rate black ratees higher than whites.[13] Teachers need to keep in mind the possibility of bias due to stereotyping in the assessment of students' performances.

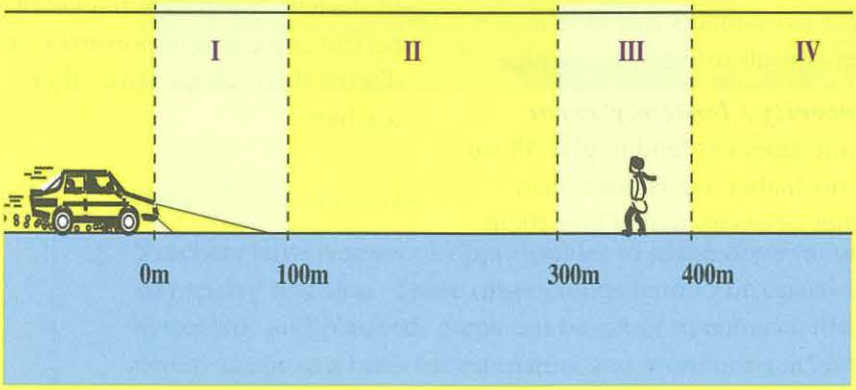
Gender stereotyping Stereotyping on the basis of gender also can influence assessments of achievement. Teachers need to be aware of the possibility that students may perform differently—or be assessed differently—depending on the mix of males and females in a group.

In colleges of further education in Scotland, lecturers wanted to maintain standards that they saw as acceptable, while at the same time giving recognition to students whom they saw as less able. The solution in some instances was to interpet 'to an acceptable standard' differently for different students.

'In building a short length of wall, a tolerance of +/- 6mm was expected of those regarded as the 'better' students, +/- 8mm for the 'average'students, and +/- 10mm for the 'less able'.'

Physics

On a clear, dark night, a car is parked on a straight, flat road. The car's headlights are on and dipped. A pedestrian standing on the road sees the car's lights. The situation is illustrated in the figure below which is divided into four sections. In which of the sections is there light? Give reasons for your answer.



Categories

5 The object reflects light and when the light reaches the eyes we see the object.

4 There are beams going back and forth between the eyes and the object. The eyes send out beams which hit the object, return and tell the eyes about it.

3 There are beams coming out from the eyes. When they hit the object we see.

2 There is a picture going from the object to the eyes. When it reaches the eyes, we see .

1 The link between eyes and object is 'taken for granted'. It is not problematic: 'you can simply see'. The necessity of light may be pointed out and an explanation of what happens within the system of sight may be given.

Students give different explanations of the relationship between light and seeing.

The 'halo' effect One of the best known systematic errors is the 'halo' effect.[13] This error arises when a teacher's assessment of one aspect of a student's performance is influenced by their assessment of other aspects of the student's performance. An example would be the tendency to rate a student's essay high on content because it also was high on language and structure.

The 'proximity' error The 'proximity error' is a tendency by assessors to give similar assessments on outcomes or criteria that are considered close together. There is a reported tendency, for example, for raters to give similar ratings on criteria that appear in

The halo effect arises when an assessor forms a general impression that causes him or her to 'fail to discriminate among conceptually distinct and potentially independent aspects of a person's behavior.'

Saal, et al.[16]

close proximity on an assessment form.[17]

The 'central tendency' error

The 'central tendency' error is a reluctance on the part of some assessors to assign very low or very high ratings.[18] These assessors are consistently more cautious in their assessments and tend to describe all students as 'average'.

'Severity / leniency' error

Some assessors tend to give all students higher assessments than other assessors would give them (the 'leniency' error[19]), while other assessors tend to give all students lower assessments than other assessors would give them (the 'severity' error). The severity/leniency error and the central tendency error usually can be overcome by providing teachers with examples of student work at each level of performance and opportunities to discuss their ratings with other teachers.

1 Victorian Certificate of Education, Studio Arts, 1994, Board of Studies, Melbourne.

2 Riverside Curriculum Assessment System (1991), Cambridge, Massachusetts: Riverside Publishing Company, p. 1.

3 The adding fractions categories were developed in discussions with Professor Tatsuoka at the University of Twente, The Netherlands, May 1992.

4 As above.

5 Mossenson, L. T., Hill, P. W. & Masters G. N. (1984) *TORCH Tests of Reading Comprehension*, Hawthorn: Australian Council for Educational Research.

6 Deluit, L. & Malcolm, C. (1994) *Using the Science Profile*, Carlton: Curriculum Corporation, pp. 64-65.

7 Deluit & Malcolm, pp. 64-65.

8 Marton, F. (1981) 'Phenomenography – A research approach to investigating different understandings of reality', *Journal of Thought*, Vol. 21, No. 3, pp. 29-49.

9 Marton, p. 31.

10 Andersson, B. & Karrqvist, C. (1981) *Light and its Qualities* (in Swedish), EKNA-rapport nr 8, Institution for praktisk pedagogik, Gothenburg University, p. 82.

11 Murphy, K. R., Balzer, W. K., Lockhart, M. C. & Eisenman, E. J. (1985) 'Effects of previous performance on evaluations of present performance', *Journal of Applied Psychology*, Vol. 70, No. 1, pp. 72-84, p. 82.

12 Landy, F. & Farr, J. (1980) 'Performance Rating', *Psychological Bulletin*, Vol. 87, pp. 77-107.

13 Kraiger, K. & Ford, J. K. (1985) 'A meta-analysis of ratee race effects on performance ratings', *Journal of Applied Psychology*, Vol. 70, No. 1, pp. 56-65.

14 Black, H., Hall, J. & Yates, J. (1988) *Assessing Modules: Staff Perceptions of Assessment for the National Certificate*, Edinburgh: The Scottish Council for Research in Education.

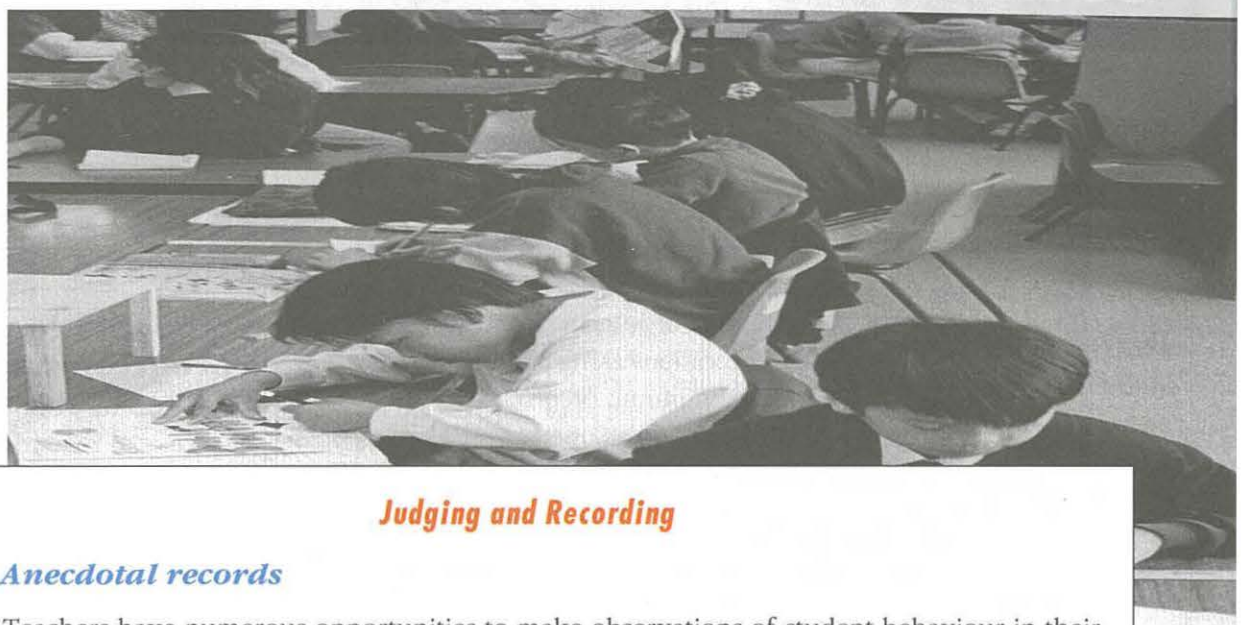
15 Thorndike, E. L. (1920) 'A constant error in psychological ratings', *Journal of Applied Psychology*, Vol 4, pp. 25-29.

16 Saal, F. E., Downey, R. G. & Lahey, M. A. (1980) 'Rating the ratings: assessing the psychometric quality of rating data', *Psychological Bulletin*, Vol. 88, No. 2. pp. 413-428.

17 Stockford, L. & Bissell, H. W. (1949) 'Factors involved in establishing a merit-rating scale', *Personnel*, Vol. 26, pp. 94-116.

18 Kingsbury, F. A. (1992) 'Analyzing ratings and training raters', *Journal of Personnel Research*, Vol. 1, pp. 377-383.

19 Kneeland, N. (1929) 'The leniency tendency in rating', *Personnel Journal*, Vol. 7, pp. 356-366.



Judging and Recording

Anecdotal records

Teachers have numerous opportunities to make observations of student behaviour in their day-to-day teaching. These observations tend to be casual and fortuitous rather than systematic and planned. Steps can be taken to enhance the usefulness of informal observations as a basis for estimating and monitoring achievement.

Rating scales

A rating scale can be used to judge and record the quality of student work. Analytic rating scales require the assessment of student work against a number of separate criteria. Holistic rating scales require a single, overall assessment of a piece of work.

Partial credit

Partial credit scoring is a method for recording the steps students successfully complete in the solution of a problem or their partially correct understandings and strategies.

Dichotomous records

In dichotomous scoring teachers assign students' performances to one of two categories such as complete/incomplete, acceptable/not acceptable, or right/wrong.

Judgements of outcomes achieved

Another method of recording is based on a judgement about whether a student has 'achieved' a particular learning outcome. The difficulty in making judgements of this kind is in knowing how much evidence is required, and in what range of contexts.

Categorising student responses

Students' responses also can be assigned to qualitatively different categories of response, possibly reflecting the different kinds of understandings that students have of a phenomenon.

Common observational 'errors'

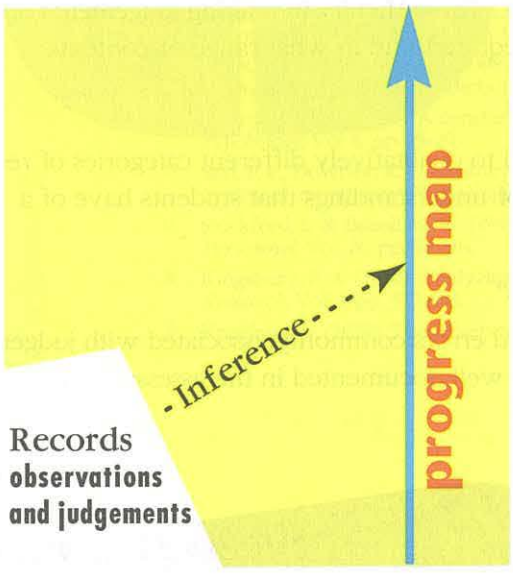
Teachers need to be aware of biases and errors commonly associated with judgements of human performance. These 'errors' are well documented in the assessment literature.

estimating attainment

The process of estimating a student's location on a progress map always involves an `on-balance' decision based on the available evidence.

The central purpose in developmental assessment is to estimate students' levels of attainment in an area of learning, conceptualised as locations along a developmental continuum or 'progress map'. To estimate attainment on a progress map it is necessary to make and record observations of students' work. When a student's level of attainment is estimated at different times during his or her schooling, it is possible to chart that student's progress over time.

The developmental continuum on page 41 has been constructed as a map for monitoring children's developing competence in spelling. Five broad levels of spelling development, labelled Preliminary Spelling to Independent Spelling, are shown. On the left of the page, one child's estimated levels of spelling attainment on four occasions are shown. Each estimate has been dated to show the child's progress as a speller (from Preliminary to Independent) over time.



making estimates

The process of estimating a student's location on a progress map is similar to the process of working out where you are on a road map from surrounding clues. In both cases observations must be matched to what is described on the map and a best estimate made from the available evidence.

Because progress maps describe paths of typical development, there is rarely, if ever, a perfect match between what is described on a progress map and observations made for any particular individual. Individuals have idiosyncratic strengths and weaknesses, and not all individuals make progress through a learning area in the same way. A strength of a progress map is that, by mapping the path of *typical* development, it provides a frame of reference for identifying and responding to the special ways in which individuals learn.

In developmental assessment, the estimate of a student's level of attainment on a progress map is based on observations and judgements of that student's work. The following discussion describes how a student's level of attainment can be estimated from four kinds of observations and judgements:

- holistic ratings
- analytic ratings
- outcomes achieved
- scored responses

Gina C Apr 1999 →

5 Independent Spelling

In this phase writers have become aware of the many patterns and rules that are characteristic of the English spelling system. When spelling a new word they use a multi-strategy approach. They have the ability to recognise when a word doesn't look right and to think of alternative spellings. Spellers in this phase will have accumulated a large bank of known words that they can automatically recall. Independent spellers continue to use personal constructions when spelling unfamiliar words in draft writing. Independent spellers realise the importance of proof reading.

4 Transitional Spelling

In this phase writers are moving away from a heavy reliance on the phonetic strategy towards the use of visual and meaning-based strategies. They may still have difficulty recognising if a word "looks right" but should be able to proof their known bank of words.

Gina C May 1998 →

Writing will show evidence of an increasing bank of words. This is a critical phase in the development of spelling. It often take writers a long time to move through it.

3 Phonetic Spelling

In this phase writers are able to provide an almost perfect match between letters and sounds. Letters are chosen on the basis of sound often without regard for conventional letter patterns. Spelling attempts are meaningful and becoming more like standard spelling. There is often evidence of self-constructed rules that may not conform to adult rules. Writers copy, recall and construct words according to their current understandings. They use rote recall for an increasing number of words.

Gina C Jun 1997 →

2 Semi-Phonetic Spelling

In this phase children show developing understanding of sound-symbol relationships. Their spelling attempts show some evidence of sound-symbol correspondence. They may represent a whole word with one, two or three letters. In this, as in all phases of development, children will be copying, recalling and inventing words. Children at this phase are able to copy letter by letter.

Gina C Apr 1996 →

1 Preliminary Spelling

In this phase children become aware that print carries a message. They experiment with writing-like symbols as they try to represent written language. Their writing is not readable by others as understandings of sound-symbol relationships have yet to develop. Children are fascinated by print and are constantly trying to explore the relationships between written and spoken words and between letters and sounds through emulating adults in role play of reading and writing.

Estimates of attainment in spelling—Gina C, April 1996 to April 1999 (First Steps Developmental Continuum).[1]

holistic ratings

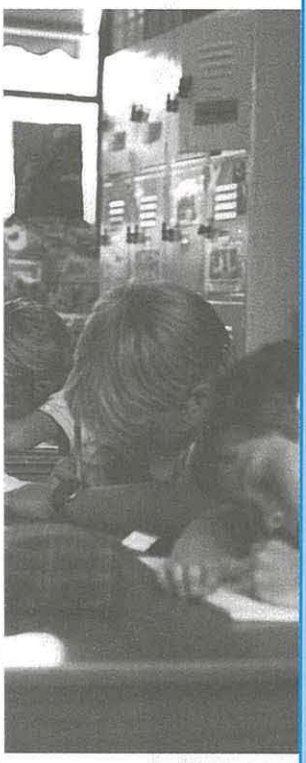
Holistic ratings are single, overall ratings of student work. Holistic ratings are commonly made of a particular piece of student work such as an essay, a project, or a piece of art work; a performance such as a dance routine, an oral presentation, or a musical performance; and a collection of work such as an art folio or a portfolio of student writing.

Holistic ratings are single, overall ratings of student work.

In making holistic ratings, teachers match the features of student work to described levels on a rating scale. Because pieces of work usually display features from more than one rating level, the process inevitably involves an 'on-balance' judgement of the most appropriate rating for each piece.

An example of a holistic rating scale is shown below. This 4-level scale was developed for use in judging the quality of students' participation in small group discussions. Teachers use the described levels to make on-balance judgements of individuals' contributions to small group discussions.

When the levels on a rating scale are constructed to correspond to



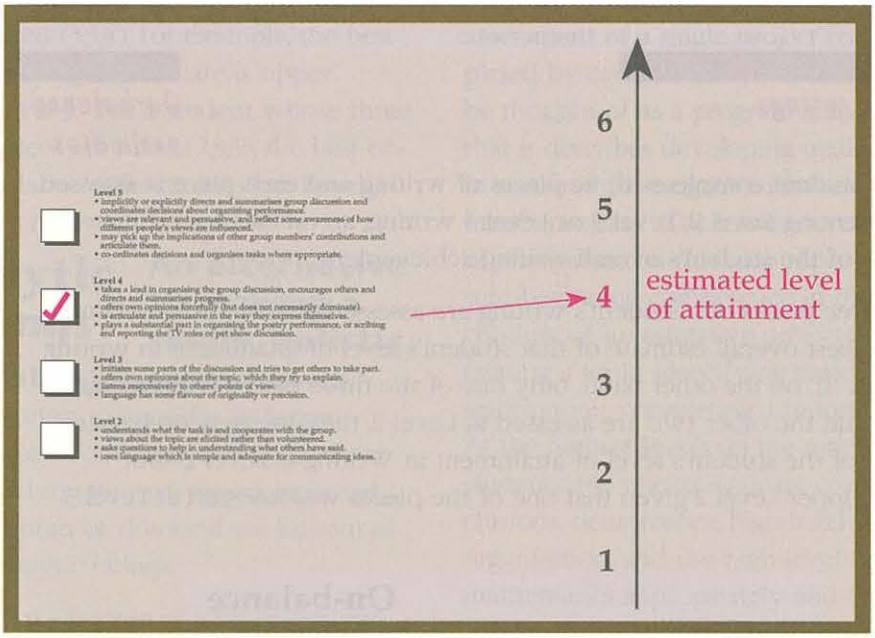
Level 5
- implicitly or explicitly directs and summarises group discussion and co-ordinates decisions about organising performance.
- views are relevant and persuasive, and reflect some awareness of how different people's views are influenced.
- may pick up the implications of other group members' contributions and articulate them.
- co-ordinates decisions and organises tasks where appropriate.

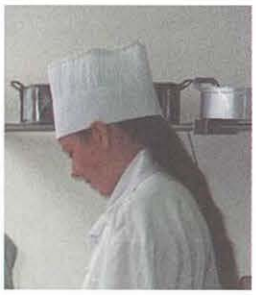
Level 4
- takes a lead in organising the group discussion, encourages others and directs and summarises progress.
- offers own opinions forcefully (but does not necessarily dominate).
- is articulate and persuasive in the way they express themselves.
- plays a substantial part in organising the poetry 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 co-operates 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.

This rating scale for making holistic judgements of a student's participation in small group discussions was constructed to correspond to Levels 2, 3, 4 and 5 of the Speaking strand of the English Profile for Australian schools (source: DART English).[2]





When the levels of a rating scale are constructed to correspond to levels on an existing progress map, the rating of a single piece of work provides a (rough) estimate of a student's level of attainment on that map.

levels on an existing progress map, each rating provides a direct estimate of a student's location on that map. The rating scale on page 42 was constructed so that each level on the scale corresponds to a level on the Speaking strand of the English Profile for Australian schools.

A student whose participation in a particular small group discussion is rated at Level 4, for example, is estimated to be performing at Level 4 of the Speaking strand of the English Profile. However, because this rating is based on only one small group discussion, it does not provide a very reliable estimate of the student's attainment in this aspect of speaking competence. A more reliable estimate could be obtained by rating the student's performance in a number of different small group discussions. The DART English kit recommends rating performance in at least three different small group discussions.

When holistic ratings are made of performances on several tasks (e.g. several small group discussions), a student's level of attainment can be estimated from these several ratings, which may be inconsistent with each other. Again, this requires an 'on-balance' estimate.

The illustration on page 44 shows how students' ratings on three separate writing tasks can be used to make on-balance estimates of their overall levels of writing attainment. If all three pieces of a student's writing are judged to be at Level 2 (2,2,2), for example, then the best estimate of that student's writing attainment is Level 2. Similarly, if all three pieces are judged to be at Level 3 (3,3,3), then the best estimate is Level 3.

On-balance estimates are required when ratings of a student's work

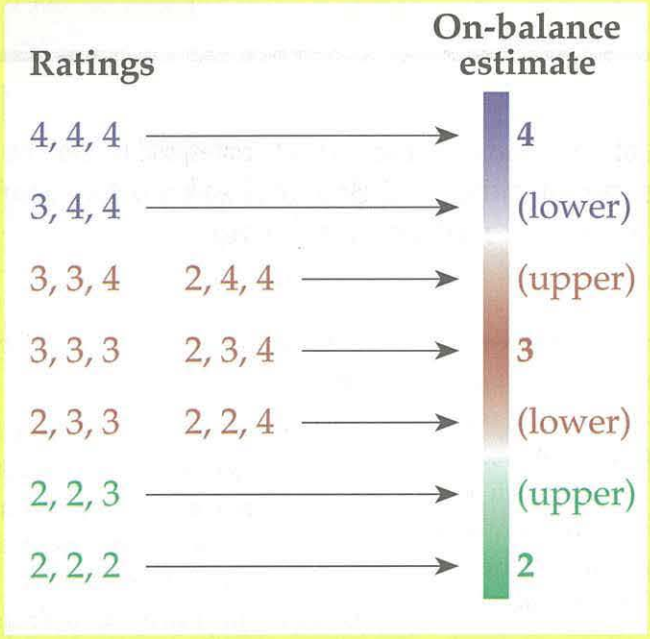
When holistic ratings of several pieces of a student's work are available, those ratings can be used to make an on-balance estimate of the student's level of attainment.

Holistic ratings

On-balance estimates

When a student completes three pieces of writing and each piece is assessed as representing Level 2, Level 3 or Level 4 writing, an on-balance estimate can be made of the student's overall writing achievement.

If all three pieces of a student's writing are assessed at Level 3, for example, then the best overall estimate of that student's level of attainment in writing is Level 3. If, on the other hand, only one of the three pieces is assessed at Level 3, and the other two are assessed at Level 2, then the best 'on-balance' estimate of the student's level of attainment in Writing is Level 2–but perhaps upper Level 2 given that one of the pieces was assessed at Level 3.



More difficult 'on-balance' decisions arise when a student displays very different levels of performance on different pieces of work. If two of a student's pieces of writing are assessed at Level 2, but the third is assessed at Level 4, should the student's Level 4 performance be treated as an 'aberration'? The strict application of an 'on-balance' decision would place this student above Level 2, and at lower Level 3.

When several pieces of student work are rated independently, an on-balance estimate of a student's level of attainment can be made.

are not identical. For a student whose three pieces of writing are rated (3,3,4), for example, the best on-balance estimate is upper Level 3. For a student whose three pieces are rated (2,3,3), the best on-balance estimate is lower Level 3.

analytic ratings

An alternative to making a single, holistic rating of a piece of student work is to consider different aspects or features of that work and to rate each aspect separately. Ratings of this kind are known as analytic ratings.

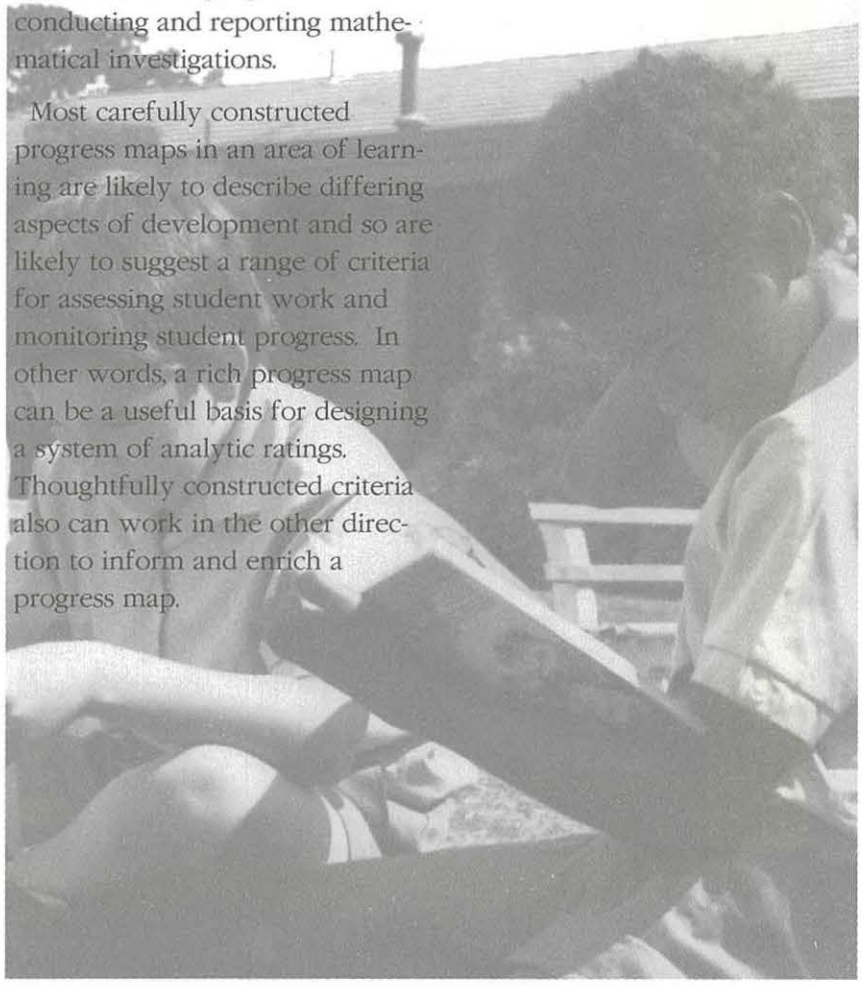
An example of a set of analytic ratings is shown on page 46. The rating criteria shown here were developed to assess students' investigative projects in a Year 12 mathematics course.[3] Teachers are asked to rate each of 18 features of a student's project as High, Medium, or Low. Further explanations and examples are provided to assist teachers in their interpretation of High, Medium and Low performances on each criterion.

Teachers are then asked to use these 18 ratings to make an on-balance decision about a student's overall level of performance on the project, expressed as a grade on a scale from E to A+ (see page 47). To assist teachers in making on-balance decisions, they are provided with descriptions of typical characteristics of projects at various grade levels (page 47) and with patterns of ratings typically associated with each grade. Page 46 shows the patterns of ratings typically associated with grades of B and B+. Very few students receive exactly these patterns of ratings and so teachers usually have to make on-balance judgements about the most appropriate grade for each project.

Although the scale from E to A+ on page 47 was developed for the assessment of a single project completed by each student, it also can be thought of as a progress map in that it describes developing quality in students' investigative mathematics projects. The lowest levels of attainment on this map represent beginning competence in the conduct of investigative projects (stating a topic, identifying basic information, completing a report). At the highest levels on the map students are able to evaluate conclusions, demonstrate high-level organisation, and use high-level mathematics appropriately and accurately. If students' mathematics projects were assessed against these criteria at various times during their schooling, then the progress map on page 47 could be used as a framework for charting over time their progress in conducting and reporting mathematical investigations.

Most carefully constructed progress maps in an area of learning are likely to describe differing aspects of development and so are likely to suggest a range of criteria for assessing student work and monitoring student progress. In other words, a rich progress map can be a useful basis for designing a system of analytic ratings. Thoughtfully constructed criteria also can work in the other direction to inform and enrich a progress map.

Analytic ratings are ratings of different aspects or features of a piece of work.



Analytic ratings

Mathematics projects

The following 18 criteria were developed for assessing projects in a mathematics course. Teachers rate projects as High, Medium or Low on each criterion. A description of each criterion and explanations of the ratings High, Medium and Low are provided to teachers.

Conducting the Investigation	High	Med	Low	Not Shown	
Identifying important information	■	■	□	□	1
Collecting appropriate information	■	■	□	□	2
Analysing information	■	■	□	□	3
Interpreting and critically evaluating results	□	■	□	□	4
Working logically	□	■	□	□	5
Breadth or depth of investigation	□	■	□	□	6

Mathematical Content	High	Med	Low	Not Shown	
Mathematical formulation or interpretation of problem, situation or issue	□	■	□	□	7
Relevance of mathematics used	■	□	□	□	8
Level of mathematics used	□	■	□	□	9
Use of mathematical language, symbols and conventions	□	■	□	□	10
Understanding, interpretation and evaluation of mathematics used	□	■	□	□	11
Accurate use of mathematics	■	■	□	□	12

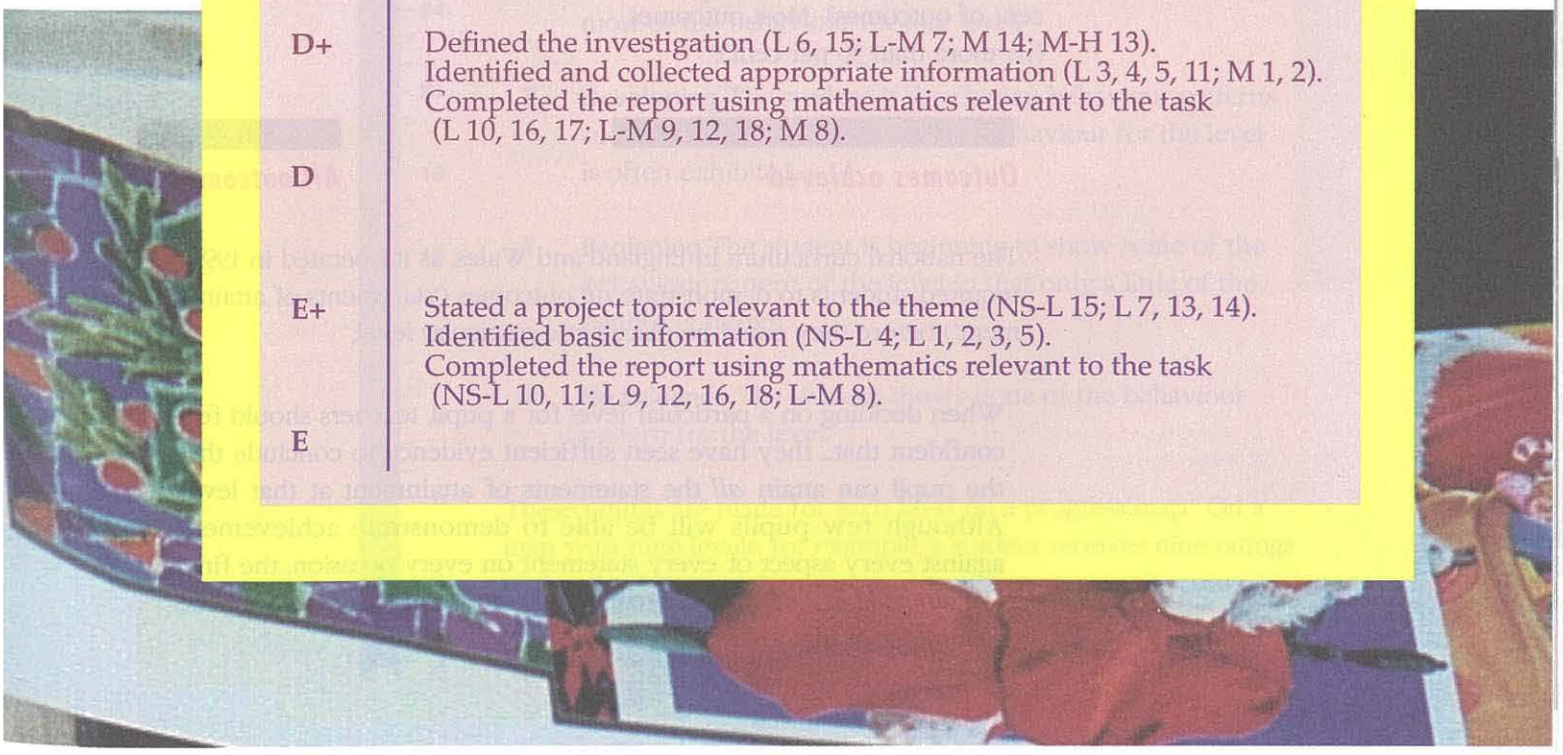
Communication	High	Med	Low	Not Shown	
Clarity of aims of project	■	□	□	□	13
Relating topic to theme	■	□	□	□	14
Defining mathematical symbols used	■	□	□	□	15
Account of investigation and conclusions	■	■	□	□	16
Evaluation of conclusions	□	■	□	□	17
Organisation of material	□	■	□	□	18

Patterns of ratings typically associated with overall grades of B (■) and B+ (■) on this mathematics project.

Analytic ratings

Progress map

A+	Clearly defined the investigation and evaluated the conclusions (M-H 6, 7, 17; H 13, 14, 15). Demonstrated high-level skills of organisation, analysis and evaluation in the conduct of the investigation (M-H 11; H 1, 2, 3, 4, 5). Used high-level mathematics appropriate to the task with accuracy (H 8, 9, 10, 12). Communicated the results succinctly in a very well organised report (H 16, 18).
A	
B+	Clearly defined the investigation (M 6, 7; H 13, 14, 15). Demonstrated skills of organisation, analysis and evaluation in the conduct of the investigation (M 4, 5, 11; M-H 1, 2, 3). Used mathematics appropriate to the task with accuracy (M 9, 10; M-H 12; H 8). Communicated the results clearly in a well organised report (M 17, 18; M-H 16).
B	
C+	Defined the investigation (L 6; M 7,15; M-H 13, 14). Demonstrated some facility in the collection and analysis of appropriate information (L 4; L-M 1; M 2, 3, 5; M 1). Used mathematics appropriate to the task (M 8, 9, 10, 12). Communicated the results in a well organised report (L 17; M 16, 18).
C	
D+	Defined the investigation (L 6, 15; L-M 7; M 14; M-H 13). Identified and collected appropriate information (L 3, 4, 5, 11; M 1, 2). Completed the report using mathematics relevant to the task (L 10, 16, 17; L-M 9, 12, 18; M 8).
D	
E+	Stated a project topic relevant to the theme (NS-L 15; L 7, 13, 14). Identified basic information (NS-L 4; L 1, 2, 3, 5). Completed the report using mathematics relevant to the task (NS-L 10, 11; L 9, 12, 16, 18; L-M 8).
E	



outcomes achieved

Records of student achievement sometimes take the form of judgements of outcomes achieved.

Records of this kind usually are based on a list of outcomes, each allocated to a particular level on a progress map. Some teachers use ticks to record each observation of an outcome and then highlight outcomes as they are judged to have been achieved.

A student who is judged to have achieved a very small number of outcomes in a level is usually considered to be 'working in' that level. Students usually are 'working in' several levels of a progress map simultaneously.

A student's current level of achievement on a progress map is sometimes interpreted as the highest level the student has 'achieved'. This raises the question of how many outcomes from a level a student must demonstrate before they can be considered to have 'achieved' that level. Must they achieve all outcomes? Eighty per cent of outcomes? Most outcomes (i.e. more than 50 per cent)?

Some assessment programs have clear rules for assigning students to a level. In the national curriculum assessment in England and Wales as it operated in 1993, for example, students had to demonstrate *all* outcomes from a level before they were judged to have achieved that level (see below).

In the First Steps program, students were first required to demonstrate *most* outcomes from a level (phase) to be assigned to that level, but later had to demonstrate all key indicators from a level (see page 49).

In other assessment programs, rules are replaced by on-balance (or best-fit) judgements. Assessors are not told what proportion of outcomes a student must demonstrate from a level, but are encouraged to make on-balance judgements of the most appropriate level given all available evidence (see page 50).

Outcomes achieved — All outcomes

The national curriculum in England and Wales, as it operated in 1993, required students to demonstrate *all* outcomes ('statements of attainment') before they could be placed at a particular level:

> When deciding on a particular level for a pupil, teachers should feel confident that... they have seen sufficient evidence to conclude that the pupil can attain *all* the statements of attainment at that level. Although few pupils will be able to demonstrate achievement against every aspect of every statement on every occasion, the final teacher assessment should broadly reflect attainment across all the statements at that level.[4]

Outcomes achieved — Key outcomes

In the early development of the First Steps continua, decisions about students' levels (phases) of development were made by recording which outcomes (indicators) were exhibited and then counting to see whether *most* indicators in a phase were demonstrated.

In order to select the most appropriate strategies to ensure that steady progress is made, it is necessary to determine the phase in which a child is currently operating. In early trials of the Continua, a child was placed in the phase in which *most* indicators were observed.

However, this was seen to be problematic because (1) most children exhibited behaviours from two or three phases; and (2) there was a question as to whether all indicators were equally significant. This led to the notion of 'key indicators':

Key indicators:

- are typical of that phase of development;
- signal the development of a significant skill or understanding;
- are exhibited by all children in that phase;
- can provide a specific focus for teaching.

Children now must exhibit *all key indicators* of a phase before they are considered to be operating in that phase.[5]

Outcomes achieved — Ratings

Patrick Griffin has proposed the following scheme for recording the extent to which students have achieved the outcomes at any given level on a progress map:[6]

3 **Established / Beyond the level** The student has established the behaviour pattern and consistently exhibits all or most of the behaviour for that level.

2 **Developing** The student is developing behaviour patterns such that some but not all of the behaviour for the level is often exhibited.

1 **Beginning** The student is beginning to show some of the behaviour pattern of the level in that only a little of the pattern is shown.

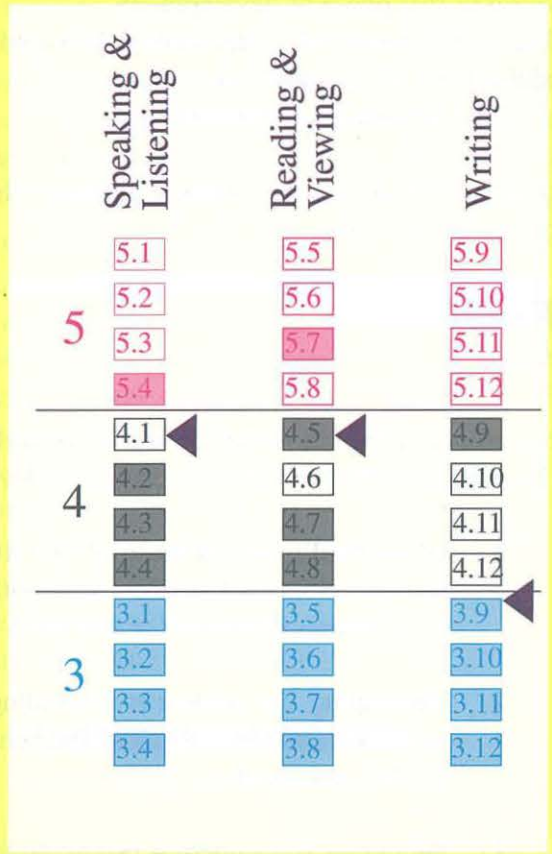
0 **No Evidence** The student shows none of the behaviour pattern for the level.

These ratings are made for each level on a progress map. On a map with nine levels, for example, a student receives nine ratings of 0 to 3. These are then summed to obtain an overall score between 9x0 = 0 and 9x3=27 for that student. A student scoring 12 is then estimated to be between Levels D and E on the map.

Outcomes achieved 'Best fit'

The Queensland Guide to Using Student Performance Standards in English encourages teachers to judge 'best fit' between students' performances and described levels of achievement.[7] This process is illustrated for several students. One of these students, Joanna, was judged to have achieved the following outcomes (shaded):

Joanna's *Speaking and Listening* performance shows greatest development in the use of skills (5.4), an aspect of language use that her teachers have emphasised. Joanna's level of performance in Speaking and Listening was determined to be more like Level 4 than any other Level.

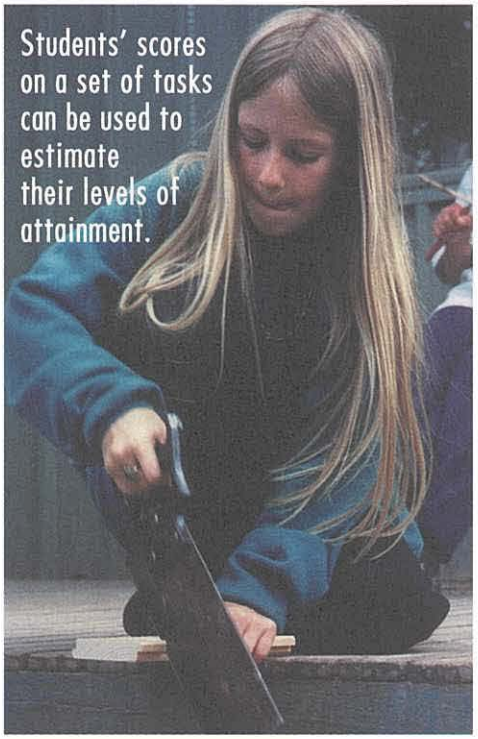


In *Reading and Viewing* Joanna's ability to interpret the features of English (5.7) is developed beyond other aspects, especially her understanding of contexts (4.6). Joanna's level of performance in Reading and Viewing was determined to be more like Level 4 than any other Level.

A smooth, typical pattern of performance in *Writing* is demonstrated. Her level of performance is below her Speaking and Reading. Joanna's level of performance in Writing was determined to be more like Level 3 than any other Level.

scored responses

Records of student performance also can take the form of scored responses to assigned tasks (e.g. classroom test questions). Students' responses to tasks can be scored either dichotomously indicating whether or not each task was completed correctly, or by assigning partial credit for partially correct solutions. If a set of tasks is assembled to address outcomes from an area of learning, then students' scored responses to those tasks can be used to estimate their levels of attainment in the area.



Scored responses

Task calibration

Performances on a set of test questions can be used to estimate students' levels of attainment on a progress map if the test questions are first developed to address the outcomes of the learning area and are then located according to difficulty (i.e. calibrated) along the map.

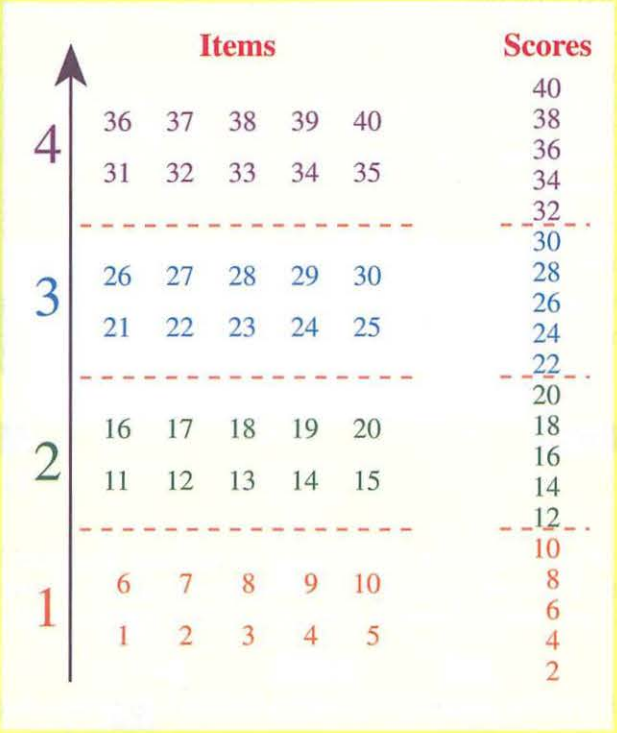
Once the area of learning to be assessed has been identified, the main steps in the process are:

- Decide on the *levels* to be covered by the assessment (an assessment at upper primary might address outcomes from Levels 3, 4 and 5 of the Australian profiles, for example).
- Develop *assessment tasks* to address outcomes from the selected levels. Each task may address a particular outcome from a particular level or may be sufficiently open-ended to allow performances at several levels.
- Use students' responses to the assigned tasks to *calibrate* all tasks along a progress map for the area of learning. The calibration process locates each task according to its difficulty (i.e. students' success rates on the task).
- *Measure* each student's attainment on this progress map. The measurement process uses the student's score on the set of tasks to estimate that student's level of attainment on the map.

Scored responses

Mapping scores

When test items are constructed to address outcomes on a progress map, students' test scores can be used to estimate their levels of attainment.



Suppose that ten items are developed to address outcomes at each of four levels on a progress map, resulting in a 40-item test. Students' scores on this test will be between 0 and 40.

Students with very low scores (between 1 and 5) on these 40 items will most probably have succeeded on only very easy (Level 1) items. A score of, say, 3 probably indicates that the student is working in Level 1. A score of 14, on the other hand, is most likely an indication that the student has achieved Level 1 and is working in Level 2; a score of 31, an indication that the student has achieved Levels 1, 2 and 3, and is beginning to work at Level 4; and so on.

When items are calibrated (positioned) along a progress map, test scores also can be positioned along the same map, allowing students' levels of attainment to be estimated from their test performances.

1 Western Australian Ministry of Education (1991) *First Steps Spelling Developmental Continuum*, Perth WA: Ministry of Education.

2 Forster, M., Mendelovits, J. & Masters, G.N. (1994) *DART Developmental Assessment Resource for Teachers*, Melbourne: Australian Council for Educational Research.

3 Victorian Board of Studies (1994) *Assessment Advice for 1994 School Assessed Common Assessment Tasks, Part 3*, Melbourne: Victorian Board of Studies.

4 The Schools Examination and Assessment Council (1993) *Pupils' Work Assessed, Geography*, London: SEAC, p. 5.

5 Western Australian Ministry of Education.

6 Griffin, P., Smith, P. & Burril, L. (1994) *The American Literacy Profile Scales*, Portsmouth, NH: Heinemann.

7 Queensland Department of Education (1994) *English in Years 1 to 10: A Guide to Using the Student Performance Standards in English*, Brisbane: Department of Education.

Estimating Attainment

Making estimates

The central purpose in developmental assessment is to estimate students' locations on a progress map from assembled evidence of their attainments.

Holistic ratings

When the rating categories of a holistic rating scale are constructed to correspond to levels on a progress map, the rating of a single piece of work provides a (rough) estimate of a student's location on that map. Holistic ratings of several pieces of work require an on-balance decision.

Analytic ratings

Analytic ratings also can be used as a basis for estimating students' levels on a progress map. On-balance estimates of levels of attainment can be made by matching students' performances to described levels on a map.

Outcomes achieved

Students' levels of attainment on a progress map can be estimated from records of the numbers (or proportions) of outcomes achieved from each level on that map.

Scored responses

Students' scores on a set of tasks can be used to estimate their levels of attainment on a progress map provided that the tasks have been 'calibrated' along that map.

reporting

Audiences for reporting:

- students
- parents
- school communities
- education systems
- the general public

The assessments teachers make and the records teachers keep often are made primarily for teaching purposes. They assist teachers to understand where individuals are in their learning and can be used to make decisions about appropriate kinds of learning experiences.

Occasionally, teachers make more formal assessments of student progress to share this information with students and parents. This sharing allows students to monitor their progress through an area of learning and to set goals for further learning. It also allows parents to inspect the progress students are making and, perhaps, to become more actively involved in their children's learning.

audiences

Students and parents are two audiences for teacher assessments. Other audiences include school communities, education systems, and the general public. These audiences usually are interested in different levels of detail about educational achievements.

Parents are likely to be most interested in knowing how children are performing in relation to teachers' expectations and in comparison with other children of the same age or grade. School communities may be most interested in knowing how children in a school are performing in relation to past levels of performance at that school or in relation to performances in other, similar schools. Education systems may be interested in monitoring the performances of particular groups of students, particularly if resources have been targeted at those groups. And the general public may be most interested in knowing that schools are teaching worthwhile knowledge and skills and that high educational standards are being maintained.

reporting against a progress map

When, teachers, schools and school systems base their assessment and reporting procedures on the principles underlying developmental assessment, reports to students, parents, school communities, system managers, and the public are likely to:

- be built around the concept of a progress map (or developmental continuum);
- provide estimates of individuals' levels of achievement on this map;
- draw on a wide range of evidence about students' achievements;
- interpret levels of achievement descriptively in terms of the kinds of knowledge, skills and understandings typical of students at each level;
- display achievements graphically, indicating individuals' or groups' estimated locations on the continuum; and
- interpret achievements by reference to the achievements of other students of the same age or grade.

Reporting against a progress map

School policies

On Bainbridge Island, Seattle, assessment, evaluation and reporting are based on some of the principles underlying developmental assessment:

The ongoing assessment of student progress, meaningful evaluation of that progress, and reporting in a manner which communicates clearly between school staff, students, and parents are critical components of successful educational programs.

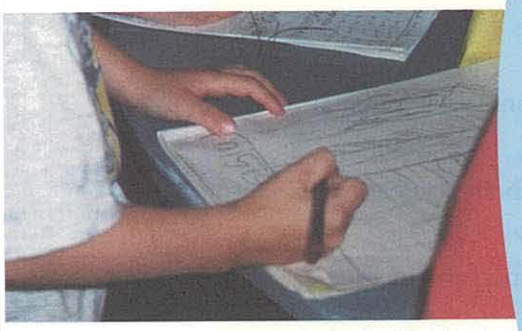
Assessment, evaluation and reporting:

- focus on what a student can do and is trying to do, based on developmental benchmarks;
- use both objective measures and professional judgements about academic performance and personal growth;
- include information about student processes, products and performances;
- document development and improvement and identify areas for growth; and
- report progress in the context of the individual learner as well as in relation to typical performance for students of the same age or grade level.

Evaluation and Reporting Handbook[1]

Many schools use progress maps in reporting student achievement to parents. Stanley Elementary School uses four stages on a developmental continuum in Reading as the basis of their report to parents of First Grade students (page 57). The stages—labelled Preconventional, Emergent, Beginning, and Developing—are accompanied by descriptions of the reading behaviours typical of students at each stage. Teachers highlight the reading behaviours each child displays in the classroom and then mark that student's estimated location on the continuum at the top of the report. By writing the date alongside each estimated location on the continuum, teachers are able to provide a general picture of a child's pattern of reading growth over time.

In the accompanying letter to parents (page 56) teachers explain that they use this continuum of reading behaviours to emphasise the ongoing nature of learning.



Dear Parents

In working with children we have come to recognize certain behaviors that all students go through as they move toward becoming readers. These behaviors can be grouped into stages along a continuum.

This progress report reflects those stages with an explanation of what behaviors are typical to that stage. The four stages on this form represent the range of reading behaviors we expect to see in first grade.

At each reporting period the classroom teacher will color in the reading behaviors your child displays in the classroom. They will also date and mark where your child falls along the continuum.

We request that you sign the report and return it to school. The report in June will be your final copy.

We are using this continuum of reading behaviors to emphasize the ongoing nature of learning. Not all the children's reading behaviors will fit neatly into one stage or another. Our goal is to provide a general picture of the pattern of growth over time.

This is the first year we are using this reporting form. We welcome your suggestions and feedback. We hope it is understandable and helpful in painting a detailed portrait of your child as a reader.

Sincerely,

Staff
Stanley Elementary Magnet School
Tacoma School District

'Our goal is to provide a general picture of the pattern of growth over time'

In this letter, which accompanies each student's report, teachers explain how the report emphasises the ongoing nature of learning.[2]

STANLEY ELEMENTARY, FIRST GRADE PROGRESS REPORT

Name

Reading Stage PRECONVENTIONAL	Reading Stage EMERGENT	Reading Stage BEGINNING	Reading Stage DEVELOPING
• Beginning concepts about print; book holding, turns pages correctly, shows start and end of book. • Recognizes own name and familiar names. • Recognizes familiar words in context (McDonalds, EXIT...). • Knows some letter names. • Focuses on pictures. • Responds to literature (smiles, claps, listens intently). • Chooses books and has favorites.	• Notices environmental print. • Begins to focus on print but uses illustrations to tell story. • Knows most letters and some sounds. • Able to memorize pattern in familiar books. • Demonstrates awareness of titles. • Can match voice to words one-to-one. • Predicts a word left out in a familiar sentence. • Participates in choral reading songs and poems. • Recognizes rhyme. • Can retell material read by adult. • Longer attention span when listening to books.	• Sees self as reader. • Relies on print more than illustrations for meaning. • Uses illustrations, sentence structure and context to read. • Recognizes familiar high frequency words in isolation. • Reads books with predictable patterns with initial prompting. • Selects own books to read. • Some awareness of author and illustrator. • Begins to use confirmation strategy. • Responds to literature through drama, art, discussion. • Can retell main ideas of text. • Can attend to book for increasing amount of time.	• Reads simple and/or predictable test. • Begins to use reading strategies for decoding. • Understands use of punctuation when reading aloud. • Begins to correct miscues of text that don't make sense. • Chooses to read a variety of books and other texts. • Able and willing to write about book. • Can retell the plot, characters, and events in simple stories. • Concentrates on reading for longer period of time.
Ages 4-6 Grade K	**Ages 5-7 Grades K, 1**	**Ages 5-8 Grades K, 1, 2**	**Ages 6-9 Grades 1, 2, 3**

Date	Classroom Teacher	Specialist Teacher	Parent / Guardian
_____	_____	_____	_____ Please sign and return
_____	_____	_____	_____ Please sign and return
_____	_____	_____	THIS IS YOUR FINAL COPY.

This first grade Reading report allows parents to monitor student progress along a reading continuum.

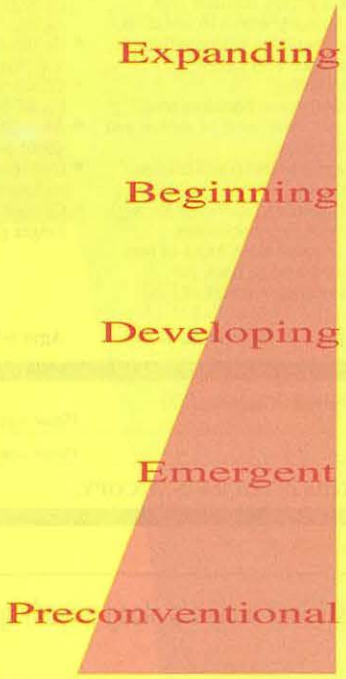
Students' achievements on the Queensland Year 6 test also are reported against a series of progress maps.[3] These maps are known as Student Performance Standards (SPS).

As for the Stanley Elementary School reading report, each continuum on the Queensland report is shown horizontally at the top of the page (see page 59). The computer-generated report shows one student's estimated levels of achievement on each of three strands of numeracy: Number, Measurement and Space. These estimated levels are marked with shaded bubbles at the top of the report.

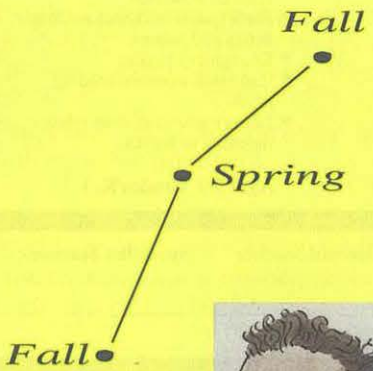
The body of the report to parents shows the items answered correctly (black circle) and incorrectly (white circle). These items are arranged in order of difficulty, with the easiest items at the bottom of the page and the hardest at the top. The report also shows which items address Level 2, Level 3 and Level 4 outcomes.

The student whose report appears on page 59 succeeded on most items in Level 2, only a few items in Level 3, and no items in Level 4. This student's results suggest that they are working in Level 2 in Measurement, Level 3 in Space, and somewhere between Levels 2 and 3 in Number.

Reporting against a progress map



Charting growth



Cindy Ruptic, a teacher of Grade 1/2 children, charts the writing progress of individual children against a developmental continuum and shares her chart with parents during parent-teacher interviews.

At the end of the first year of using the continuum, Cindy Ruptic graphed the reading and writing development of her first and second graders on a chart. She used a dot on the left to indicate each child's stage of writing development in the fall. In the spring, she made a second dot to indicate each child's stage of writing development and drew a line to connect the two dots. Her chart clearly showed the growth her students made in five months. This technique provides Cindy with a graphic picture of reading and writing growth in her classroom that she can share with parents on Back-to-School night.[4]

Aspects of Numeracy

● shows this student's result.

Number	Measurement	Space
● *This student answered this item correctly.* ○ *This student did not answer this item correctly.*		
Level 4		
○ Represent common fractions on a number line. ○ Use place value to compare and order numbers.	○ Convert measurements using common metric prefixes. ○ Use conventional units of mass. ○ Calculate areas of rectangles.	○ Visualise position and describe it on a map using distance, direction and co-ordinates. ○ Make use of conventions relating to co-ordinate pairs. ○ Visualise locations and understand directional language when reading maps.
Level 3		
○ Completely solve division problems by interpreting remainders. ○ Divide a whole number by a 1-digit number. ○ Subtract a 3-digit number from another involving regrouping. ○ Recognise equivalent fractions. ● Continue number patterns. ● Place whole numbers in order. ○ Subtract one 3-digit number from another. ○ Partly solve division problems by interpreting remainders. ○ Multiply by a 1-digit number. ○ Interpret whole numbers written in words and use a calculator for adding whole numbers. ● Represent word problems as number sentences. ○ Use a calculator for subtracting whole numbers.	○ Use a calculator to add lengths expressed as decimals. ● Compare areas by counting units. ○ Use an object as a repeated unit of measurement. ○ Identify a right angle. ○ Calculate time intervals. ○ Choose shapes that can cover a region with no gaps or overlaps. ○ Compare and measure length to the nearest graduation.	○ Visualise and follow paths using co-ordinates. ● Recognise features of a 3D object that are shown in a 2D diagram. ○ Recognise the same shape within arrangements and patterns. ○ Recognise 3D shapes from a description of their surface. ● Select a flat shape that will fold to make a prism.
Level 2		
○ Continue whole number patterns involving addition. ● Add 3-digit whole numbers. ● Multiply small whole numbers. ○ Subtract small whole numbers. ● Add 2-digit whole numbers. ● Recognise place value in whole numbers.	○ Recognise that different units can be used to measure the same length. ○ Locate a date on a calendar. ● Read time on a clock. ● Read a thermometer scale to the nearest marked graduation.	● Compare properties of 3D shapes. ● Interpret placement of objects in drawings.

Students' results on the Queensland Year 6 numeracy test are reported in relation to three levels on Number, Measurement and Space progress maps.

reporting against typical progress

It is possible to show parents how a student is achieving in relation to other students of the same age or grade.

When schools and school systems report student achievement against a progress map, parents are provided with information about the kinds of knowledge and skills students have displayed and indications (usually graphical) of students' estimated levels of achievement on a continuum. By mapping the achievements of a number of students it is possible to show parents how an individual is achieving in relation to other students of the same age or grade.

The Queensland Year 6 report (page 59) provides not only details of each student's performances on all test items and an estimate of that student's location on each of the Number, Measurement and Space strands, but also an indication of how that student performed in relation to all other Year 6 Queensland students. The student whose report is shown on page 59 is estimated to be achieving in the bottom 20 per cent of Year 6 students on each of the Number, Measurement and Space strands.

A 'growth curve' shows rates of growth over time.

To show parents how a student is achieving in relation to other students it is necessary to collect data on the performances of a wider group of students: perhaps all students, or a representative sample of students, at a particular age or grade within an education system.

When a progress map is constructed as a framework for monitoring progress over a number of years of school, it is possible to show on the map the locations of students of a particular age or in a particular grade. In this way, it is possible to conclude that an individual is achieving at the same level as the average Year 3 student, for example, or at a level above 90 per cent of Year 6 students.

The progress map on page 61 shows arithmetic tasks covering a wide range of mathematical development. These tasks have been 'calibrated' on the map according to difficulty, with low-level tasks towards the bottom of the map, and higher-level tasks towards the top. The locations of the tasks are based on the performances of nationally representative samples of students across a range of grade levels.[5]

The performances of students in each grade between Grades 1 and 10 have been used to construct bands showing the kinds of tasks mastered by the average student in the grade during that particular year. The average third grade student, for example, was not able to answer questions such as 76-12=? at the beginning of the third grade, but could answer such questions by the end of the year.

It can be seen from the picture on page 61 that the greatest growth in arithmetic skills occurs in the first three years of school. Much less growth occurs in the seventh, eighth and ninth years of school, perhaps because mathematics instruction has by then moved on to give greater attention to other areas of mathematics learning such as algebra, and chance and data.

This decline in the rate of growth in arithmetic skills is shown in a slightly different way by the 'growth curve' on page 62. This curve shows students' average levels of achievement on this progress map after every two or three months of school. The steepest growth occurs in the first few years of school. The growth curve becomes less steep in the upper primary and especially lower secondary school.

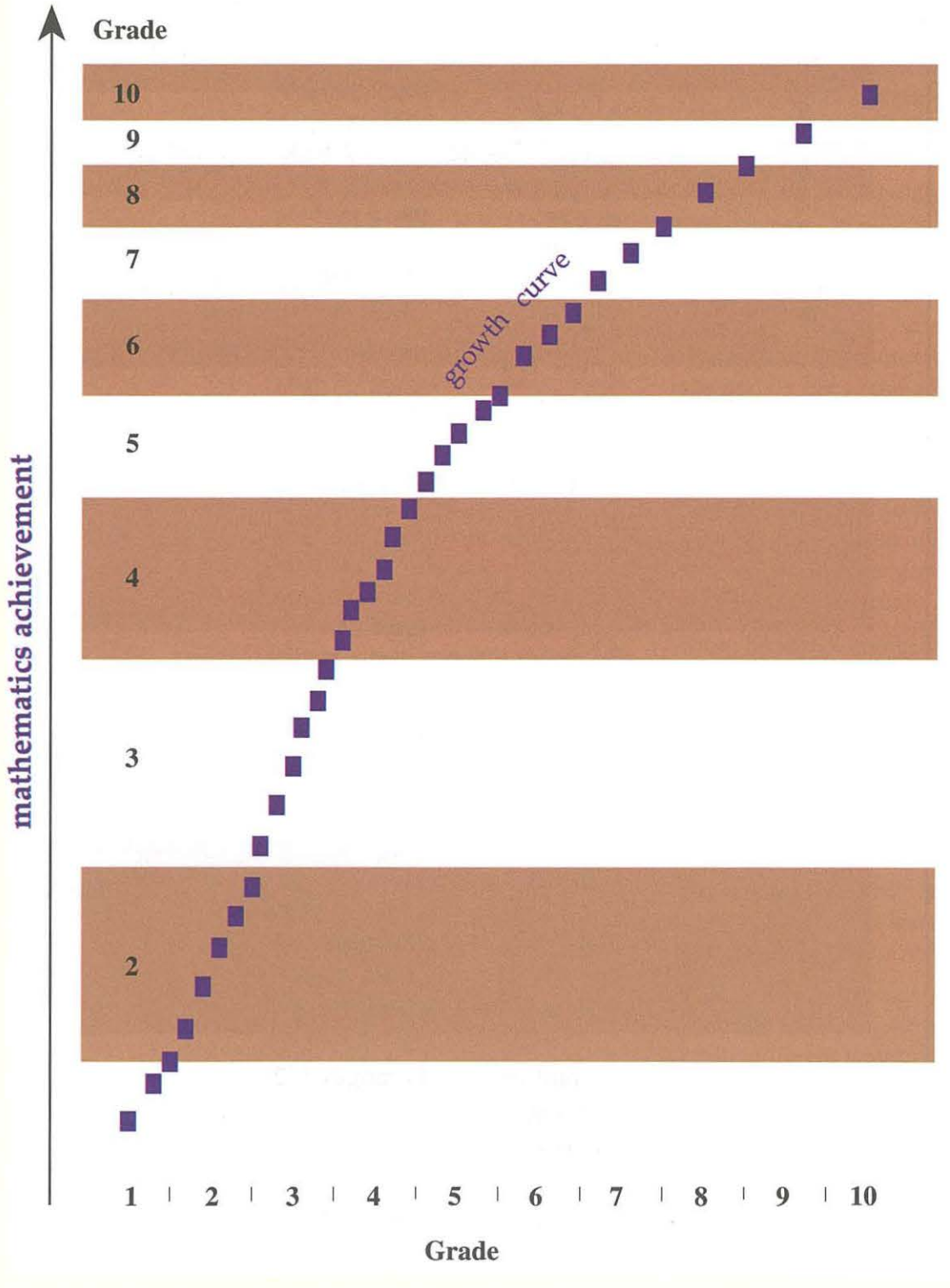
Arithmetic map

Arithmetic tasks calibrated on a progress map. Tasks in each band are examples of tasks that the average student in that grade probably could *not* complete correctly at the beginning of the year, but probably could do by the end of the year.[6]

mathematics achievement

Grade			
10			
9	$5\frac{1}{2} \times 4$ $3/8 \div 1/2$	$6\frac{1}{4} - 2\frac{2}{3}$ $5\frac{7}{8} - 3\frac{1}{4}$	
8	\$100.00 - \$99.95 $3\frac{1}{4} + 2\frac{1}{8}$	$5\frac{1}{2} + 2\frac{5}{8}$ \$24.60 ÷ 3	1308 ÷ 12
7	75 x 75 7 x \$30.40	105 ÷ 15	
6	14 x 25 132 ÷ 4 \$409.74 + \$93.25 2/4 + 1/4	62.07 - 7.9 8 x 75 500-94	69 ÷ 3
5	370-82 6 x 15	27.3 + 24.09	
4	25-16 15 ÷ 5 3 x 8 86+29 2 x 4	2391+548+1210 6 ÷ 2 14-6	
3	76-12 66+4 7+9 47+2	3 'fours'	
2	2 'threes' 8-2 5+4 5-3 1+3	8 oranges ÷ 4	
1	8-4 buttons 2+4 frogs 3+2 birds 5-2 buttons	4 oranges ÷ 2 1+2 matches	

Progress in arithmetic also can be shown by a 'growth curve' which is steep in the early years of school but shows less growth in arithmetic skills in the upper primary and lower secondary school. A progress map is a prerequisite for graphing and monitoring growth.



An alternative to graphing only average levels of achievement at particular ages or grades on a progress map is to indicate the *distribution of achievements* at each age or grade. This approach has been taken by Peter Hill and Ken Rowe who have plotted the achievements not only of the 'average' student in each Year level, but also of the bottom 10%, bottom 25%, top 25%, and top 10% of achievers in each Year group (see page 64).[7]

The growth curves on page 64 are plotted against the levels of the Victorian literacy profiles, in this case in Reading. Once again, the steepest growth occurs in the earliest years of school and progress in reading becomes slower in the upper primary school and especially in the lower secondary school.

One interesting feature of Hill and Rowe's graph of reading progress is the much more rapid and sustained progress of the top 10% of readers (see growth curve for the 90th percentile) than for the bottom 10% of readers (10th percentile) over the compulsory years of schooling. The gap between the best and worst readers in any cohort increases throughout the years of school.

Another interesting feature is the large spread of reading achievement within any one Year group, and the considerable overlap across Year levels. Fifty per cent of Year 3 students have reading levels above the lowest achieving 25 per cent of Year 5 students; 10 per cent of junior secondary school students still read at the level of the average Year 3 student.

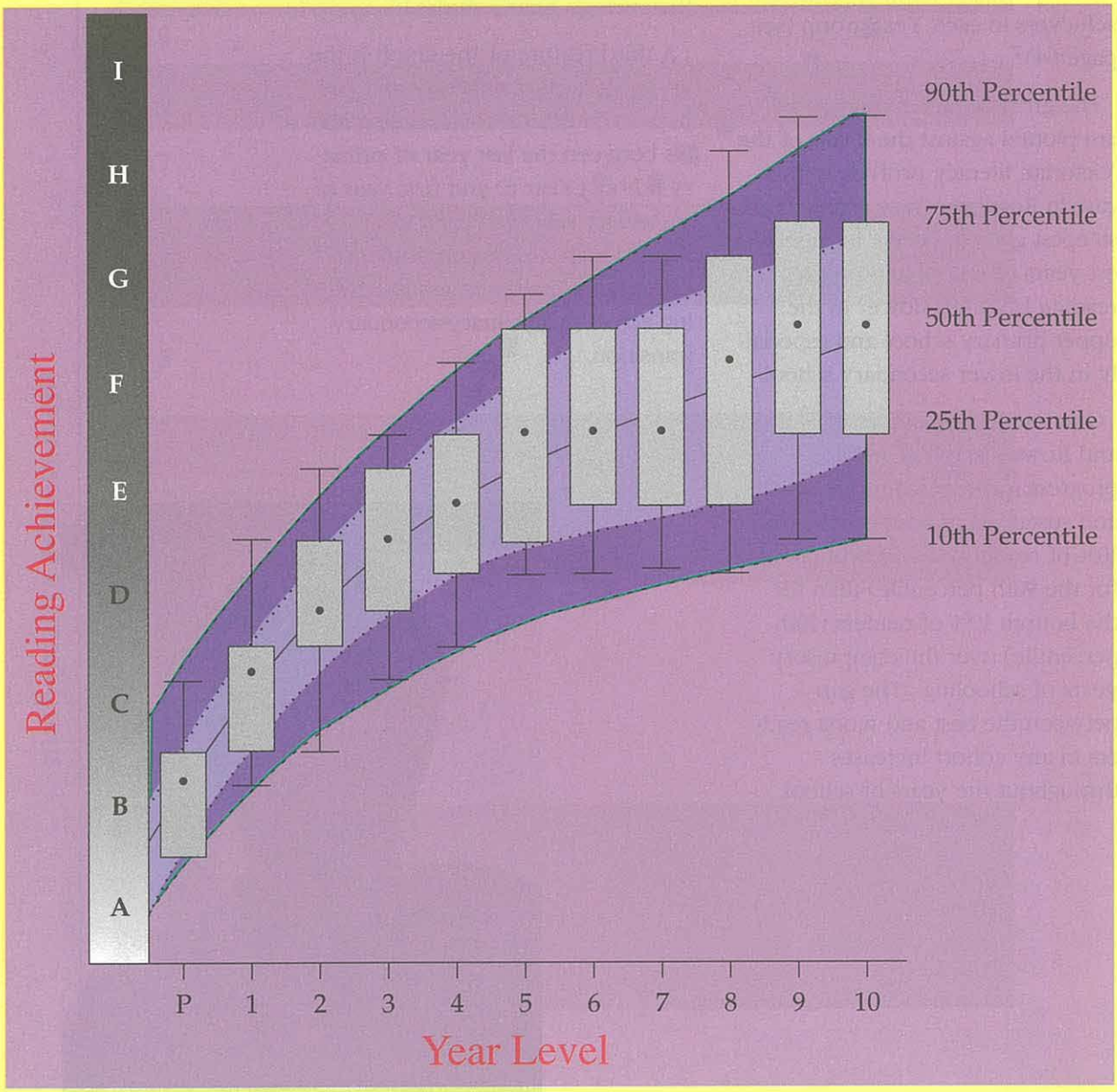
A third feature of the graph is the apparent lack of increase—and perhaps even decrease—in reading levels between the last year of primary school (Year 6) and first year of secondary school (Year 7). According to this graph, students make no overall progress in reading across this primary-secondary transition.

The gap between the best and worst readers in any cohort increases throughout the years of school.

Reporting against typical progress

Growth in reading

Peter Hill and Ken Rowe have graphed the reading achievements of students from the beginning of school to Year 10 against the progress map provided by the Victorian literacy profiles (also known as the American Literacy Profile Scales[8]). Their graph shows growth curves in reading for students at the 10th, 25th, 50th, 75th and 90th percentiles.



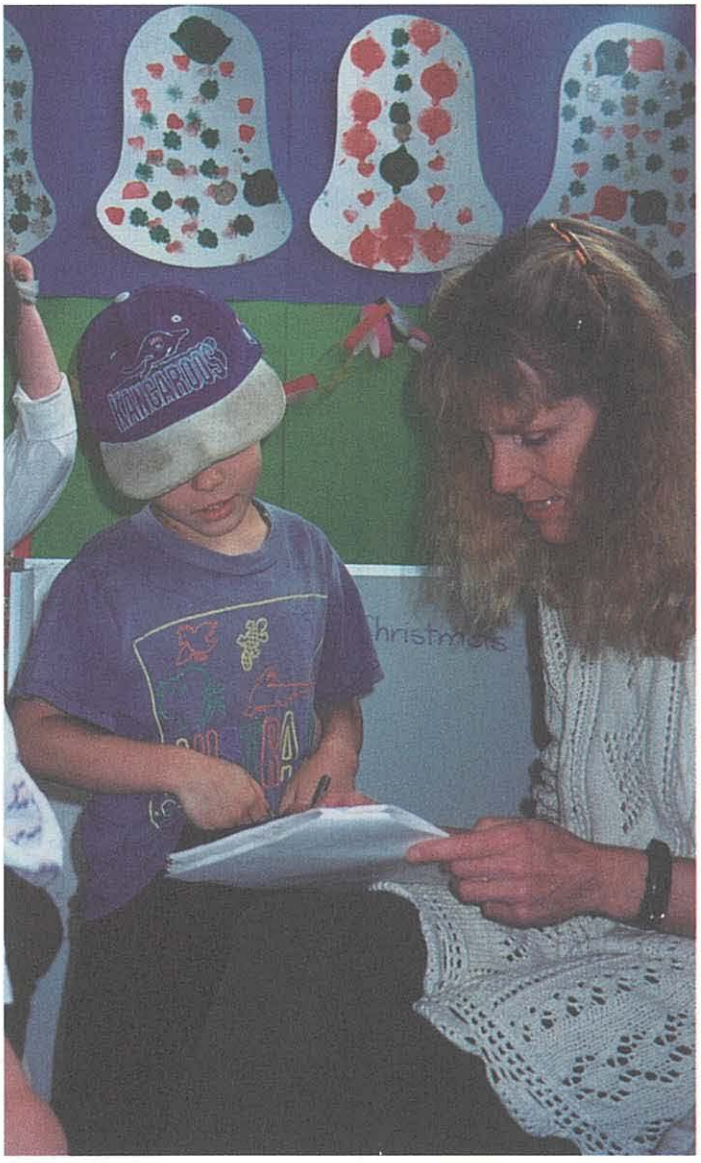
monitoring standards

A progress map also provides a framework against which the performances of particular groups of students can be plotted and compared, and changes in levels of achievement can be monitored over time.

The US National Assessment of Educational Progress, for example, has monitored the reading achievements of nationally representative samples of students against a progress map in reading for more than 20 years.[9] Positions along this reading continuum are indicated by numbers in the range 0 to 500. The reading behaviours typically displayed at five levels along the continuum—labelled Rudimentary (150), Basic (200), Intermediate (250), Adept (300), and Advanced (300)—have been described and illustrated (see page 66).

The graph on page 67 shows trends in the reading achievements of four groups of students (white 9-year-olds, black 9-year-olds, white 17-year-olds, and black 17-year-olds) over the period 1971 to 1988. It can be seen from this graph that, throughout this period, the average reading level of white 9-year-olds was above the Basic level of reading achievement and below Intermediate. The average reading levels of white 17-year-olds remained around the Adept level of achievement.

In the same period, the average reading levels of black 9-year-olds increased from near Rudimentary to near Basic, and the average reading levels of black 17-year-olds increased from near Intermediate to near Adept. These marked changes in average reading levels among black students were monitored and reported against the reading levels shown here in a series of national reports.



Monitoring standards **Reading levels**

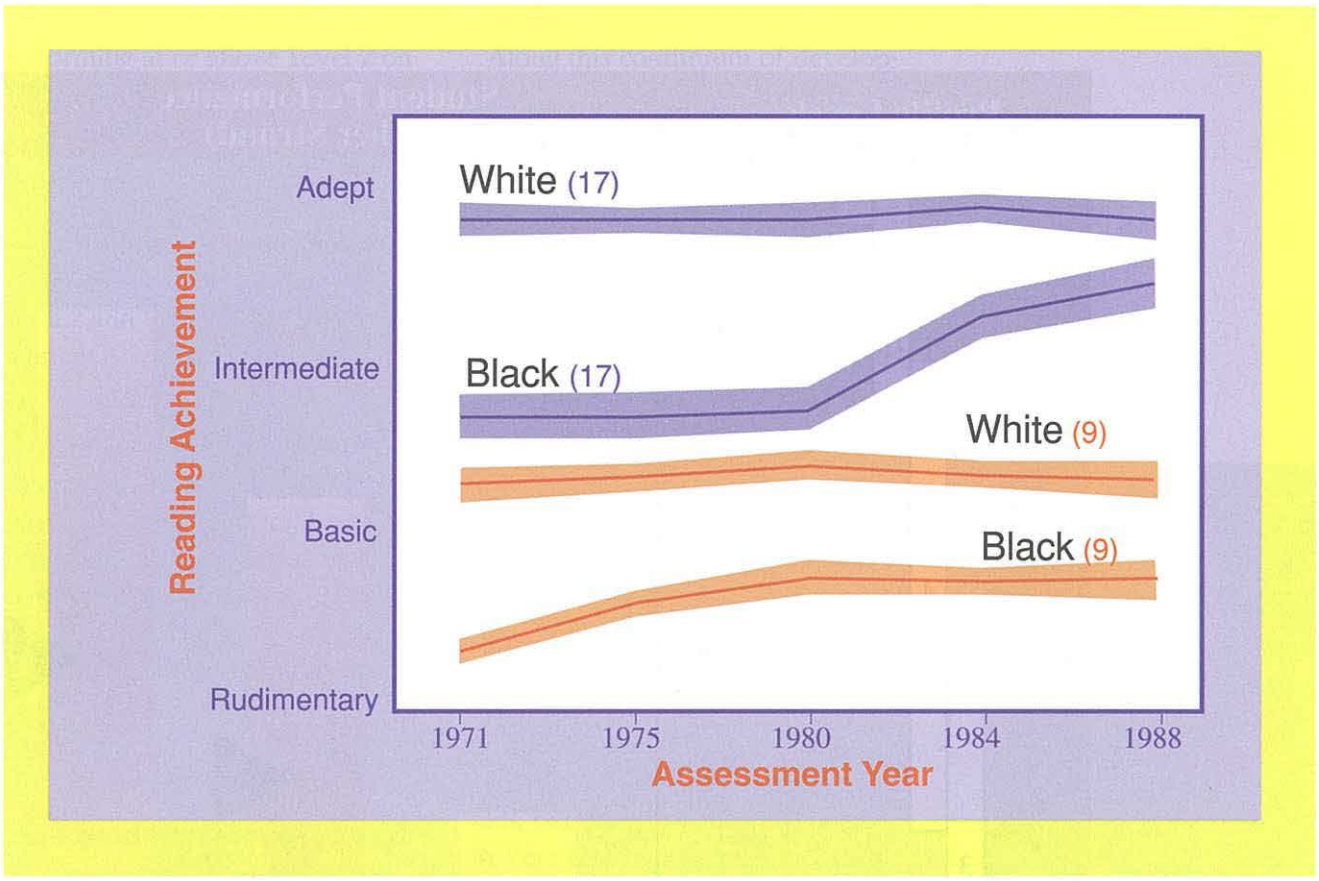
350 **Advanced**
Readers who use advanced reading skills and strategies can extend and restructure the ideas presented in specialized and complex texts. Examples include scientific materials, literary essays, historical documents, and materials similar to those found in professional and technical working environments. They are also able to understand the links between ideas even when those links are not explicitly stated and to make appropriate generalizations even when the texts lack clear introductions or explanations. *Performance at this level suggests the ability to synthesize and learn from specialized reading materials.*

300 **Adept**
Readers with adept reading comprehension skills and strategies can understand complicated literary and informational passages, including materials about topics they study at school. They can also analyze and integrate less familiar material and provide reactions to and explanations of the text as a whole. *Performance at this level suggests the ability to find, understand, summarize, and explain relatively complicated information.*

250 **Intermediate**
Readers with the ability to use intermediate skills and strategies can search for, locate, and organize the information they find in relatively lengthy passages and can recognize paraphrases of what they have read. They can also make inferences and reach generalizations about main ideas and author's purpose from passages dealing with literature, science, and social studies. *Performance at this level suggests the ability to search for specific information, interrelate ideas, and make generalizations.*

200 **Basic**
Readers who have learned basic comprehension skills and strategies can locate and identify facts from simple informational paragraphs, stories, and news articles. In addition, they can combine ideas and make inferences based on short, uncomplicated passages. *Performance at this level suggests the ability to understand specific or sequentially related information.*

150 **Rudimentary**
Readers who have acquired rudimentary reading skills and strategies can follow brief written directions. They can also select words, phrases, or sentences to describe a simple picture and can interpret simple written clues to identify a common object. *Performance at this level suggests the ability to carry out simple, discrete reading tasks.*

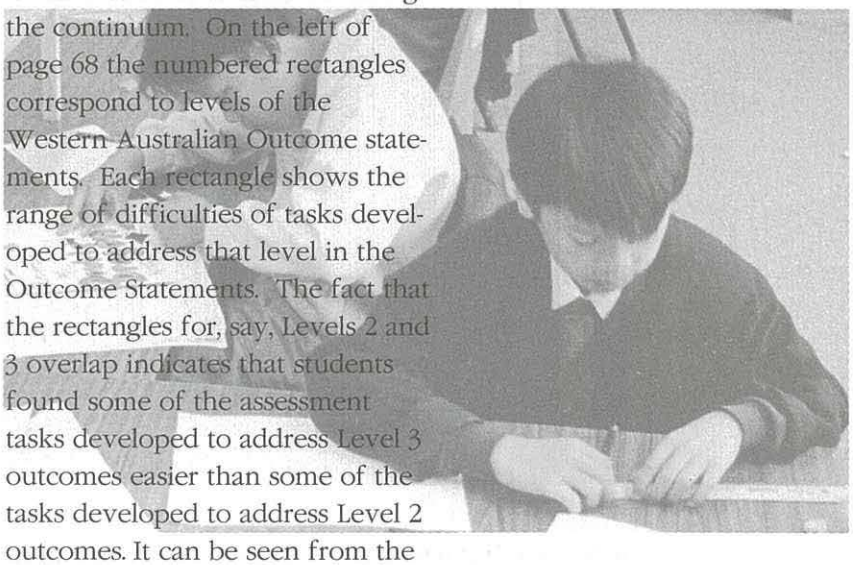


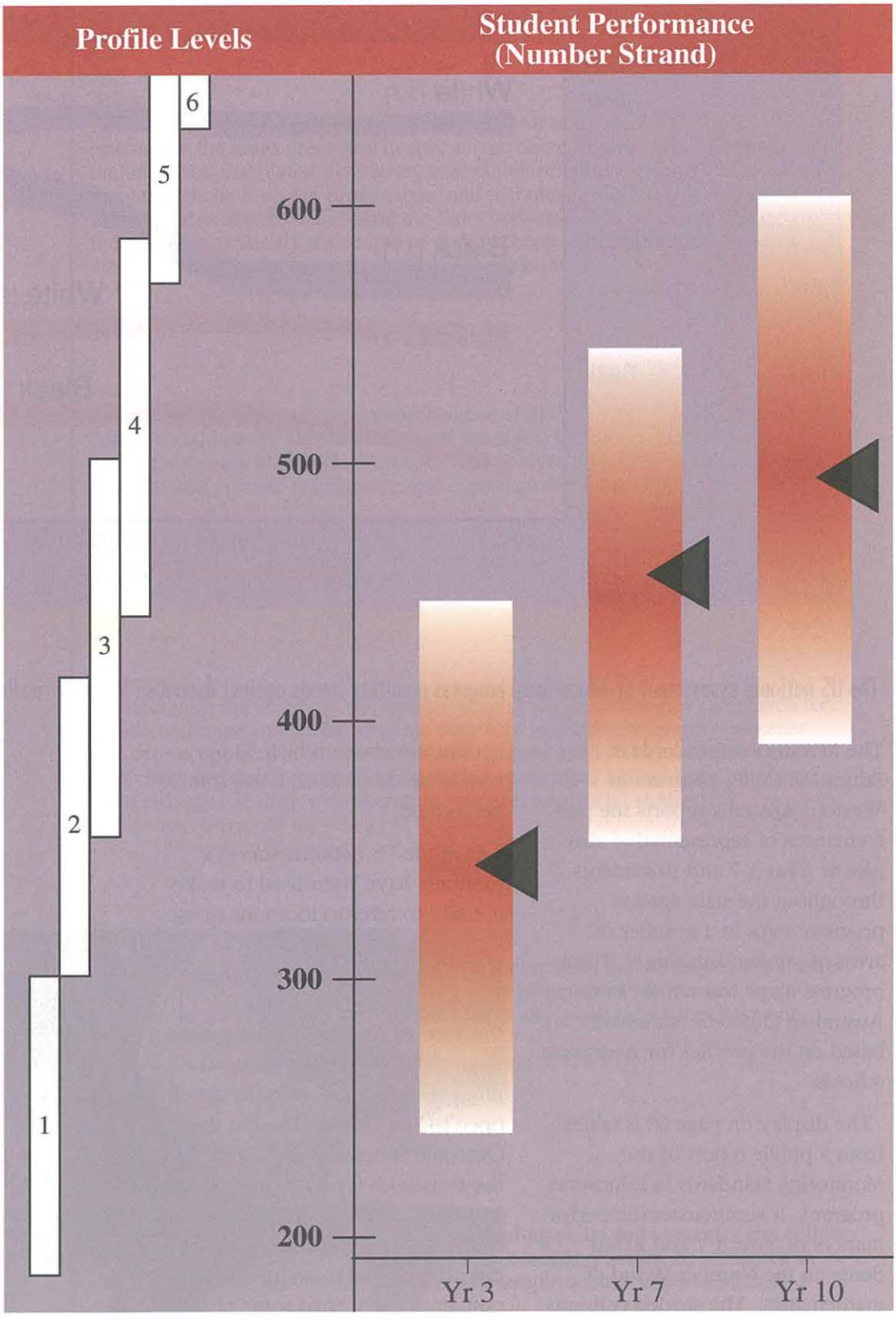
The US national assessment of educational progress monitors trends against described levels of reading achievement.

The Monitoring Standards in Education (MSE) program in Western Australia reports the performances of representative samples of Year 3, 7 and 10 students throughout the state against progress maps in a number of areas of student learning.[10] These progress maps, known as Western Australian Outcome Statements, are based on the profiles for Australian schools.

The display on page 68 is taken from a public report of the Monitoring Standards in Education program. It summarises the performances of Year 3, 7 and 10 students on the Number strand of mathematics. The shaded columns mark the ranges of the Number strand within which 80 per cent of each age sample performed. The triangle shows the level on the continuum above which 50 per cent of students in each age sample performed.

As in the US national surveys, numbers have been used to make it easier to refer to locations along the continuum. On the left of page 68 the numbered rectangles correspond to levels of the Western Australian Outcome statements. Each rectangle shows the range of difficulties of tasks developed to address that level in the Outcome Statements. The fact that the rectangles for, say, Levels 2 and 3 overlap indicates that students found some of the assessment tasks developed to address Level 3 outcomes easier than some of the tasks developed to address Level 2 outcomes. It can be seen from the display on page 68 that 50 per cent of Year 3 students were





In Western Australia, educational standards are monitored against progress maps in various areas of the curriculum. Triangles show the levels achieved by 50% of students in Years 3, 7 and 10. Much greater growth in Number occurs between Years 3 and 7 than between Years 7 and 10.

performing at or above Level 2 on these Number tasks. The kinds of knowledge, skills and understandings typically displayed by these students can be seen by referring to the descriptions and examples of behaviours associated with Level 2 and lower Level 3 in Number.

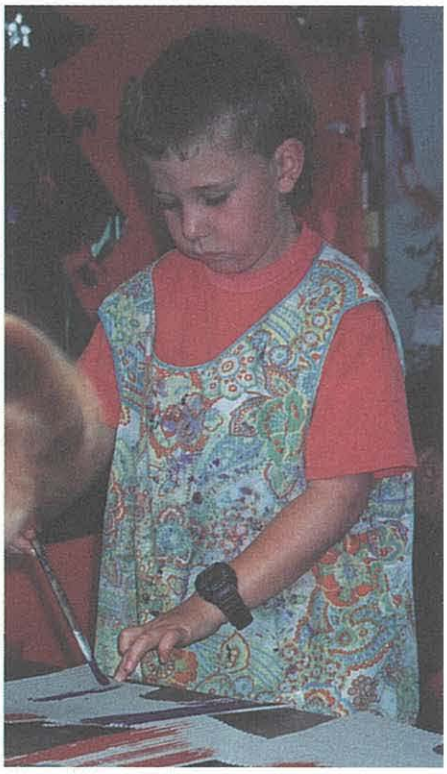
To show the achievements of Year 3, 7 and 10 students on the same progress map in Number, it was necessary to first 'calibrate' all assessment tasks on the same map. This was done by ensuring that some of the more difficult tasks given to younger students were also given to older students, thus providing a 'link' between the three assessment instruments.

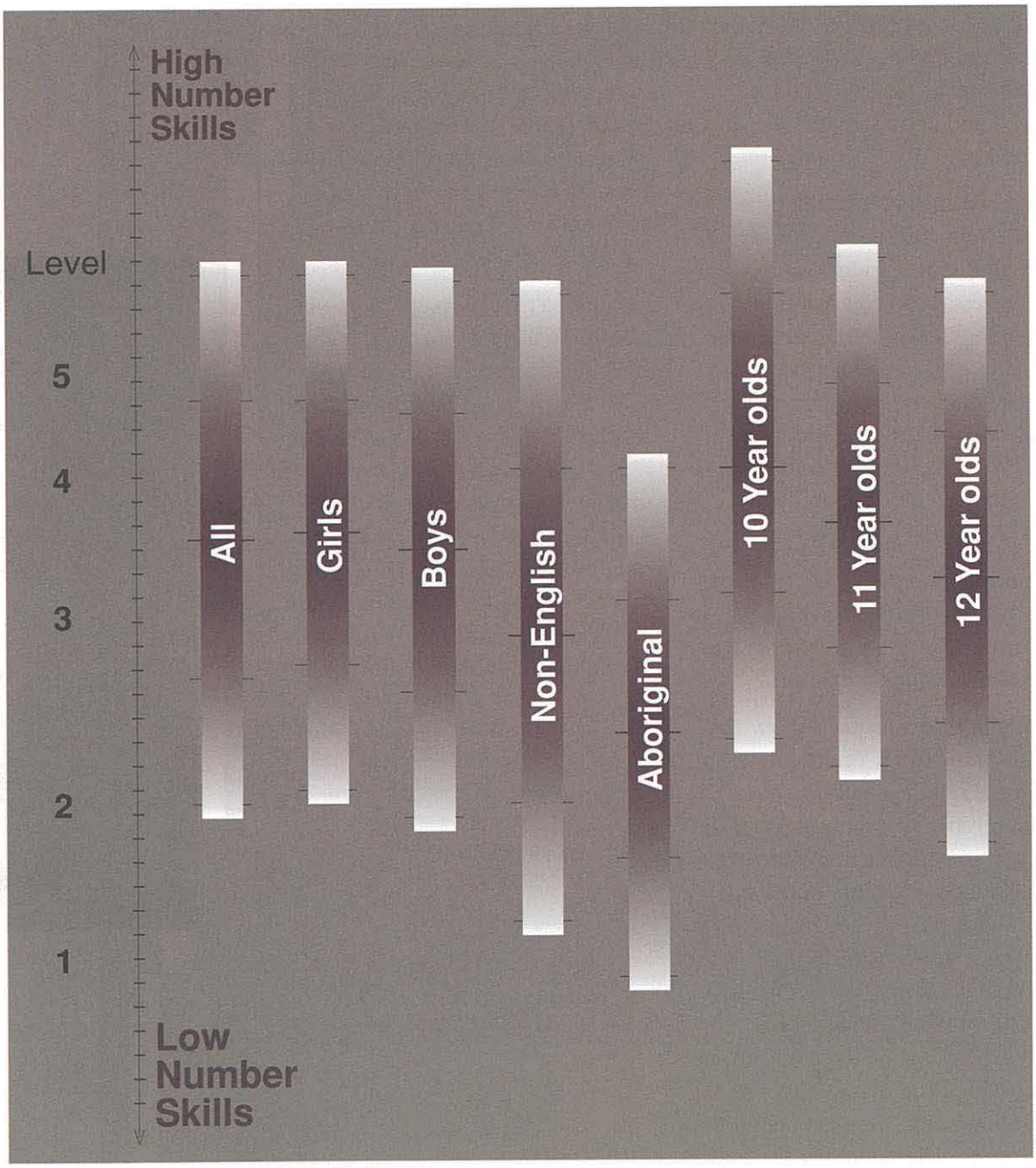
A progress map also can be used as a framework for studying the performances of various subgroups within a student population.

The graphs on page 70 show the distributions of all Year 6 students in New South Wales, boys, girls, non-English speaking background students, Aboriginal and Torres Strait Islander students, and 10-, 11- and 12-year-olds in Year 6 against a progress map for the Number strand of mathematics.[11] The shaded column for each subgroup shows the range of the continuum along which the middle 90 per cent of students were estimated to be located. The range for the middle 50 per cent of students is also shown.

Along this continuum of developing Number skills, five levels of achievement were identified. The Number skills typically demonstrated by students at each of these five levels of achievement were described and illustrated in the public report of this statewide assessment program, allowing readers of the report to see the kinds of Number skills being achieved by various student groups.

In the picture on page 70, 10-year-olds have higher estimated levels of achievement than 11-year-olds, who in turn have higher estimated levels of achievement than 12-year-olds. The reason for this is that the groups shown here are not representative of all 10-, 11- and 12-year-olds, but of 10-, 11- and 12-year-olds in Year 6. Most students in Year 6 are 11; only the more advanced 10-year-olds are likely to be in Year 6.





In the New South Wales basic skills tests, the performances of particular subgroups of the student population are plotted against the levels of described progress maps in aspects of literacy and numeracy.

Reporting

Audiences

Students and parents are two audiences for reporting. Other audiences include school communities, managers of education systems and the general public.

Reporting against a progress map

In developmental assessment, levels of achievement are reported in terms of a progress map. Levels of student achievement usually are displayed graphically and interpreted descriptively.

Reporting against typical progress

When schools and school systems report achievement against a progress map, students' performances can be interpreted not only in terms of the knowledge, skills and understandings described on that map, but also in terms of the performances of other students of the same age or grade.

Monitoring standards

A progress map provides a frame of reference against which the achievements of groups of students can be plotted and compared, and in terms of which educational standards can be monitored over time.

1 Bainbridge Island School District (1993) *Evaluation and Reporting Handbook Grades Kindergarten - Fifth*, Bainbridge Island: School District #303, p.2.

2 The Stanley Elementary Magnet School first grade report and accompanying letter to parents was provided by Professor Bonnie Campbell-Hill.

3 Queensland Department of Education (1993) *Student Performance Standards: English*, Brisbane: Queensland Department of Education.

4 Campbell Hill, B. & Ruptic, C. (1994) *Practical Aspects of Authentic Assessment: Putting the Pieces Together*, Norwood, MA.: Christopher-Gordon Publishers Inc, p. 242.

5 The report for parents reproduced on page 59 was distributed to parents of all students sitting the Queensland Year 6 Test in literacy and numeracy in 1995.

6 *Keymath Diagnostic Arithmetic Test* by Austin Connolly, William Nachtman, and Milo Pritchett © 1972 American Guidance Service, Inc. 4201 Woodland Road, Circle Pines, MN 55014-1796 USA. Reproduced with permission of the Publisher. All rights reserved.

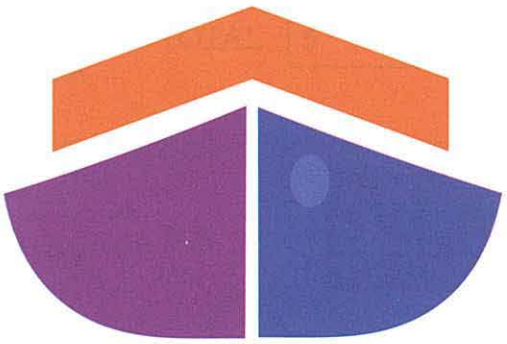
7 Rowe, K., Hill, P. & Holmes-Smith, P. (1994) *Assessing, Recording and Reporting Students' Educational Progress: The Case for Profiles.* Paper presented at the annual conference of the Australian Association for Research in Education, Newcastle.

8 Griffin, P., Smith, P. & Burril, L. (1994) *The American Literacy Profile Scales*, Portsmouth, NH: Heinemann.

9 Educational Testing Service (1990) *The Reading Report Card: National Assessment of Educational Progress*, Princeton: ETS.

10 Titmanis, P., Murphy, F., Cook, J., Brady, K. & Brown, M. (1993) *Profiles of Student Achievement: English and Mathematics in Western Australian Schools*, Perth, WA: Ministry of Education.

11 Masters, G., Lokan, J., Doig, B., Khoo, S-T., Lindsey, J., Robinson, L. & Zammit, S. (1990) *Profiles of Learning: The Basic Skills Testing Program in New South Wales*, Melbourne: Australian Council for Educational Research.



Please photocopy and fax or mail.

ASSESSMENT RESOURCE KIT

ORDER FORM

Australian Council for Educational Research
347 Camberwell Road (Private Bag 55)
Camberwell Victoria AUSTRALIA 3124
Telephone: (03) 9835 7447 Fax: (03) 9835 7499
International: Tel: 61 3 9835 7447 Fax: 61 3 9835 7499
Email: sales@acer.edu.au
Order online: www.acerpress.com.au

Charge to

Name *or*
Organisation _____

Purchase Order no. *(if applicable)* _____

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	A100ARK	AUD$9.90		
Portfolios	A101ARK	AUD$9.90		
Performances	A102ARK	AUD$9.90		
Projects	A103ARK	AUD$9.90		
Progress Maps	A105ARK	AUD$9.90		
Products	A104ARK	AUD$9.90		
Assessment Methods	A106ARK	AUD$9.90		
Paper and Pen	A107ARK	AUD$9.90		
Educational Measurement	A108ARK	AUD$9.90		
Videotape: Understanding Developmental Assessment	A700ARK	AUD$54.95		
Implementing Developmental Assessment: Workshop Manual & Videotape	A500ARK	AUD$38.50		

ADD FREIGHT:
Within Australia: 10% of invoice value, min. $5.50 max. $27.50 (inclusive of GST).

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

SUBTOTAL
Freight
TOTAL $

☐ Please add my name and address to the mailing list for further information about ARK materials when they become available.

CHARGE: ☐ Account ☐ Cheque Enclosed (AUD$)

☐ Bankcard ☐ Mastercard ☐ American Express ☐ Visa ☐ Diners Club

☐☐☐☐ ☐☐☐☐ ☐☐☐☐ ☐☐☐☐

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

ABN 004 398 145

Developmental Assessment 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