Assessing, Recording and Reporting Students' Educational Progress: The Case for Profiles

Kenneth J. Rowe, Peter W. Hill
Centre for Applied Educational Research
The University of Melbourne

and

Philip Holmes-Smith
Strategic Planning and Policy Division
Directorate of School Education, Victoria


Acknowledgments

The willing assistance of the schools and teachers who contributed to the data is gratefully acknowledged, as is the administrative and financial support of the Directorate of School Education, Victoria.

Correspondence should be addressed to: Ken Rowe, Centre for Applied Educational Research, The University of Melbourne, John Smyth Building, Parkville, Victoria 3052, Australia. [Tel: (03) 344 8741; Fax: (03) 347 0945]
Assessing, Recording and Reporting Students' Educational Progress:  
The Case for Profiles

Abstract:  In response to persistent concerns about limitations in the exclusive use of traditional testing programs for student assessment and standards monitoring, there is now a widespread movement in many OECD countries for developing more "in context" and "authentic" approaches to educational assessment.  For example, Newmann & Archbald (1990) argue that "...most data currently used to assess schools' performance, especially scores on standardised tests, fail to measure meaningful forms of human competence and that significantly new forms of assessment need to be developed" (p. 164).  One approach to addressing these concerns is the development of Subject Profiles as frameworks for assessing, recording and reporting student progress - an approach in which Australia has taken a leading role.  This paper defines what is meant by the term Subject Profiles, provides a brief account of their historical origins, and indicates why and how Subject Profiles have been developed.  In arguing the case for the use of Profiles in both monitoring and explanatory educational research, particular attention is given to their practical utility at the individual student level, as well as at the class, school and system levels.  To this end the paper presents data from recent studies for 34,000 students in Prep. to Year 10, drawn from 650 government, Catholic, independent primary and secondary schools, using the three strands of reading, writing, and spoken language from the English Profiles (Victoria, 1991) and the two strands of space and number from the Mathematics Profiles (Victoria, 1992).

The Context

Current educational policy activity related to accountability, assessment, standards monitoring, performance indicators, quality assurance and school effectiveness is widespread, both within Australia and elsewhere.  At the core of this activity, and central to any consideration of the effectiveness of educational provision are issues concerned with the assessment, recording and reporting of student learning outcomes.  Whereas the long-term goals of school education may be expressed as the enhancement of young peoples' access to and participation in society, as well as preparation for meeting the constantly changing demands of the modern workplace (eg., Dawkins, 1988, 1990; OECD, 1986), "...the immediate and most direct measures of outcome in schools are measures of student progress" (Masters, 1994:2).  Masters continues: "Ultimately, most of what we do in school education -- including our efforts to improve administrative structures and the quality of the teaching-learning environment -- can be judged in terms of their implications for enhanced student learning" (p. 2).  Herein, however, lies a dilemma that Nisbet (1993:25) highlights as follows:
In today's schools, assessment is a main influence on how pupils learn and how teachers teach. Whether assessment is in the form of examinations and tests, or marks and grades for coursework, its influence is pervasive. Often it distorts the process of learning through teaching to the test, cramming, short-term memorising, anxiety and stress -- to the extent that learning to cope with assessment has become almost as important as the genuine learning which such assessments are supposed to measure. For many young people, assessment dominates education.

To date, for the purposes of standards monitoring and accountability, the measurement of learning outcomes at the student, school, system or national levels has relied almost exclusively on the use of standardized achievement tests. While the use of such tests for the measurement and identification of educational effectiveness is typically justified on the grounds of maximising reliability, it is often argued that this has been at the expense of validity (see Moss, 1994) and has been the focus of intense discussion and criticism (see Broadfoot, 1986; 1988). For example, in summarising the British, European and North American attempts at curriculum and assessment reform during the 1970's, Lacey and Lawton (1981:229-230) note:

...conventional standardized achievement tests have inherent risks as instruments of evaluation for accountability since they seldom cover more than the common core or very basic curriculum units. Thus, as the sole instrument, they may be highly deceptive because of lacking content validity. ...test scores as such have low information value about the outlying processes as well as the environmental and administrative frame conditions necessary to understand and appreciate the skills and efforts needed to fulfil a certain educational goal.

Similarly, within the expanding school effectiveness research literature (see Bosker, Creemers & Scheerens, 1994), the identification of 'effective schools' continues to be made on the basis of limited operational definitions of what it means to be a 'good school'. The most common method is to choose those schools with aggregated scores on standardized tests of reading and mathematics (or on public examinations), that are higher than average, after making adjustments for measures of student intake characteristics. Newmann & Archbald (1990) argue that "...most data currently used to assess schools' performance, especially scores on standardised tests, fail to measure meaningful forms of human competence and that significantly new forms of assessment need to be developed" (p. 164).

Recent International and Australian Developments

In response to such concerns, many large education systems, particularly those in Europe, Britain, Canada, New Zealand and the
United States that have had student monitoring procedures in place for many years, have recently embarked on the process of dismantling and re-building their assessment programs to make them more inclusive of the broad range of competencies required of students, and to improve their utility in terms of reporting to students themselves, parents, employers and to the community. Such reforms include: the California Assessment Program (1991), the developments recommended by the Task Group on Assessment and Testing in England and Wales (TGAT, 1988), the Policy on Curriculum and Assessment in Scotland (1990), and the Toronto Board of Education's Benchmarks - Standards of Student Achievement (see Rutledge, 1993). All these reforms are illustrative of attempts to develop more "in context" (Gardner, 1992) and "authentic" approaches to assessment (Archbald & Newman, 1992; Hill, 1993; Newman & Archbald, 1992). Key features of these new systems include:

- a concern for better methods of assessing and reporting on the actual and developing achievements and competencies of individual students,
- the development of national or system-wide standards frameworks for assessment and reporting,
- the use of a broad range of assessment methods, including 'common' or 'standard' tasks undertaken by students, and
- a significantly enhanced role for teacher observation and judgement in relation to these tasks, and the assessment of competency.

These criterion-referenced and 'standards-based' approaches to assessment focus on tasks that involve "in-context" skills, as well as reflecting student-centred approaches to teaching and learning. For example, Nisbet (1993) uses the term "authentic assessment" to refer to "...methods of assessment which influence teaching and learning positively in ways which contribute to realising educational objectives, requiring realistic (or 'authentic') tasks to be performed and focusing on relevant content and skills, essentially similar to the tasks involved in the regular learning processes in the classroom" (p. 35). In a similar vein, Shavelson (1994) uses the term "performance assessment" to refer to:

...concrete, well-contextualized tasks or activities that are sampled from a subject-matter domain...and that often call for an integration of conceptual understanding...to carry them out...a performance assessment is scored by combining both the conclusions drawn from an investigation and the procedures used to arrive at those conclusions (p. 235).

Commenting on the impact of this re-building process related to recent developments in performance assessment techniques, Shavelson (1994:235-236) notes:

At this writing, performance assessments as measures of educational
achievement are front and centre on international and U.S. assessment agendas. The Third International Mathematics and Science Study is committed to including performance assessments as part of its comparisons of educational achievement. The U.S.'s National Assessment of Educational Progress has mandated by law to include performance assessments in its national surveys of educational achievement.

A central feature of this effort is the emphasis placed on the interdependence of assessment and instruction. For example, Nisbet (1993) notes: "There is a strong movement in many countries...for a radical new approach to assessment..." in which "...the distinction between assessment and instruction is removed; assessment is not an artificially separate event but an integral part of instruction (p. 34). Despite some criticism of this view (eg., Messick, 1994), the movement is gathering momentum in the USA, in Australia, in Britain and in many OECD countries.

In a review of these developments from their early stages, Masters (1991) noted that while much can be learned from experiences and developments in other places, it is important to recognise that Australia is already at the forefront of major efforts to design and introduce new approaches to the description, assessment, measurement and monitoring of student achievement. In fact, Australia has technical expertise and experience that few countries can match and have provided models of practice that have attracted attention worldwide. Moreover, the recent introduction into some Year 12 systems of common assessment tasks requiring teacher observation and professional judgement has placed Australia several years ahead of other countries whose student performance measures rely solely on traditional, standardised pencil and paper tests (see Masters & Hill, 1988; Hill, Brown & Masters, 1993).

The Development of Profiles in Constructing a Standards Framework

One approach to standards monitoring that runs parallel to these reforms in assessment is the development of Profiles as frameworks for recording, reporting and monitoring students' educational progress - an approach in which Australia has taken a leading role. Following recommendations for a national focus on assessment and standards monitoring within Australia (Dawkins, 1988), work began in 1990 on a project under the auspices of the Australian Education Council (AEC) and its Curriculum Assessment Committee (CURASS), designed to provide useful national frameworks in the form of Profiles for the recording and reporting of student achievement, against which the performances of individuals or groups could be assessed and monitored. The reports of the AEC Review Committees (Finn, 1991; Mayer, 1992) illustrate this focus. For example, Finn (1991:57-58) recommended that:

All States/Territories develop valid and publicly credible, nationally
comparable ways of assessing young people's attainments in terms of the National Competence Profile Levels in the areas of: (1) Language and Communication, (2) Mathematics, (3) Problem Solving, (4) Science and Technological Understanding, (5) Cultural Understanding, and (6) Personal and Interpersonal.

In the context of recommending that a Standards Framework for these Key Competencies be developed, the Finn report stated: The most difficult step in the process of making the competencies operational is the development of each of them into a useable profile within a consistent framework describing different levels of achievement against which progress can be assessed and reported (p. 59). Before, during and since this time, however, the term Profile has had wide currency throughout the educational community and continues to be used and interpreted in a multitude of ways. As a consequence, there is considerable confusion surrounding the term, as well as the ways it has been applied to its related products.

Conceptions of Profiles. Baumgart (1986) notes that the term Profile has long been used in the literature of educational measurement and evaluation ...to describe the presentation of an individual's achievements or characteristics on multiple dimensions" (p. 42). In recent years, the term has taken on a particular meaning related to the reporting of students' academic and other achievements including personal qualities in the form of specific, desegregated statements or checklists of students' achievements and attributes. For example, Mansell (1986), one of the earliest advocates of Profiles, defines "profiles" or "records of achievement" as "...documents constructed by professional teachers or trainers...describing as accurately and succinctly as possible the knowledge, skills and experience of an individual relative to a particular curriculum" (p. 25).

A broader definition of a Profile is provided by the U.K.'s AMMA (1983) as: "...a systematic, comprehensive description and assessment of a pupil's academic and non-academic achievements, attributes and interests, set out in a format easy to interpret both by educational and non-educational users." (p. 2). Reynolds and Wheatley (1988) provide a generic definition: "A profile is defined as a document or record of student achievement and profiling as the process by which a profile is generated" (p. vii). More recently, the fundamental idea underlying the Profiles developed by CURASS in Australia during the period 1990-1993 was Glaser's (1981) notion of a "progression of increasing competence" (p. 935), such that Boomer (1992), the first chair of CURASS, describes a Profile as "...a vertical map of performance territory from lower to higher performance" (p. 63).

Two Approaches to the Development of Profiles. Common to each of these definitions and their exemplification is an emphasis on first specifying what students are expected to do and then checking to see whether they can. This has led to two major approaches for the development of Profiles.
The first approach is what Masters (1994) refers to as the `can do', 'checklist' or 'outcomes-based' model that has been strongly influenced by the behavioural objectives, mastery learning and criterion-referenced testing movements of the 1960's and 1970's (eg., Glaser, 1963; Popham, 1978). Characteristic of this first approach is the development of broadly-specified lists of observable, mostly decontextualised student outcomes, using verbs such as list, read, write, count, identify, state, calculate, etc. For example, one of the desired outcomes in Band A for the Mathematical Inquiry strand of A National Statement on Mathematics for Australian Schools (1991), states: "Clarify, use and interpret mathematical terms and phrases" (p. 41). While this approach constitutes a valuable start towards more explicit specifications of the kinds of knowledge, skills and understandings we might wish to see students develop, it is highly unlikely that "...meaningful can/cannot judgements can be made about broad outcomes of this kind" (Masters, 1994, p. 6). Even with more explicit specifications of outcome statements and delineation of contexts, Jessup (1991), Masters (1994), Nuttall & Goldstein (1986) and Wolf (1991) among others have warned that the key difficulties confronting the checklist or 'can do' model is that: (1) the outcome statements are subject to wide variability in interpretation, (2) are often too poorly defined to ensure comparability, and (3) are unlikely to provide a reliable basis for monitoring student performance standards over time.

The second approach to the development of Profiles, and characteristic of the related work done in Australia on curriculum-specific Subject Profiles since 1988, is what Masters (1994) refers to as "achievement maps":

This approach, like the first, seeks to provide a more explicit identification of outcomes and a framework against which...the progress of an individual, a school, or an entire education system can be mapped and followed. But this approach is built not around the notion of an outcomes checklist, but around the concept of growth...Student progress is conceptualised and measured on a growth continuum, not as the achievement of another outcome on a checklist (Masters, 1994, p. 9).

Locations along this 'growth continuum' are illustrated by descriptive indicators of stages or 'levels' of increasing competence typically displayed by students at those locations. Unlike the 'can/cannot do', checklist model which describes student achievement deterministically, the levels of achievement along the growth continuum are described and reported probabilistically. That is, rather than attempting to make unequivocal 'can do' judgements, this 'growth model' approach to the development of Subject Profiles aims to provide estimates of a student's current and developing level of achievement and,

...provides an accompanying description of the kinds of understandings
and skills typically displayed by students at that level. To estimate a student's current level of achievement, a wide variety of assessment instruments can be used (Masters, 1994, p. 11).

This last point needs to be stressed. Despite numerous appeals and prevailing wishes to the contrary (e.g., Stevenson, 1983), it is important to note that Subject Profiles (including those based on the checklist model) are not assessment instruments per se. Rather, they provide recording and reporting frameworks for assessment. Unfortunately, this point is not well understood. In reporting the development of the reading and writing bands of the Victorian Literacy Profiles project that began in 1987, Griffin and Nix (1991:191) stress:

It is important to emphasise that the reading and writing band scales are not assessment instruments. They are means of communicating an interpretation of observable behaviours. This means that appropriate assessment tasks (such as the TORCH test, assignments and other assessment tasks) need to be identified and their relationship to the reading and writing bands needs to be defined.

Similarly, Boomer (1992:63) notes:

The subject profile is not itself an instrument of assessment, something which does not appear to be well understood. It is, if you like, a vertical map of performance territory from lower to higher performance upon which a student's cumulative performance as assessed can be placed.

More recently, in outlining the ways in which the national profiles might be used, Hill (1994:38) writes:

In making use of the profiles it is assumed that a wide range of both formal and informal assessment methods could be used in arriving at professional judgements regarding students' achievement levels, including direct observations of student performance, written assignments, class-based and standardised tests, specially constructed assessment tasks, and so on. The results of these assessments can then be reported using the common language of the national profiles. In other words, the profiles function as a framework for assessment and reporting and do not in themselves constitute an assessment method. What they do allow, however, is for teachers, schools and school systems to communicate about student progress and achievement using a language and standards which are consistent across classrooms, schools and school systems.

Hill further notes: "For the profiles to function effectively as an assessment and standards monitoring framework, it is necessary that typical patterns of student performance in relation to the eight levels be established empirically" (p. 38). This is a crucial point on two
grounds. First, with the exception of the Literacy, English and Mathematics Profiles (Victoria, 1990, 1991, 1992) and the Western Australian Student Outcome Statements (see Rowe, 1994), national, system or sample estimates of student performance levels on Subject Profiles have yet to be obtained. Second, again with the exception of the Victorian Profiles mentioned above, the constituent descriptive indicators or 'pointers' of the Subject Profiles (and their related 'bands' or 'levels') that have been developed to date in Australia (and elsewhere) have not been empirically calibrated. Rather, the constituent indicators have mostly been determined by curriculum experts (albeit in consultation with teachers), with the performance levels being set a priori and often arbitrarily. In our view, the implications of this lack of empirical validation for the major elements of Subject Profiles or subject-based Student Outcome Statements are problematic. That is, we would contend that in the absence of such validation, much of the value of Subject Profiles will be eroded and could bring their use as reliable frameworks for monitoring student achievement standards into disrepute. Thus, the key purpose of this paper is to define and illustrate what is meant by the term Subject Profile and to indicate how and why they have been developed in Victoria. In arguing the case for the use of Profiles in both monitoring and explanatory educational research, particular attention is given to their practical utility at the individual, class, school and system levels. To this we present data from several studies for 34,000 students in Prep. to Year 10, drawn from 650 government, Catholic, independent primary and secondary schools, using the three strands of Reading, Writing, and Spoken Language from the English Profiles (Victoria, 1991) and the two strands of Space and Number from the Mathematics Profiles (Victoria, 1992).

What is a Subject Profile?

The term Subject Profile, as used in this paper, refers to a shared framework of empirically calibrated descriptive performance indicators, located on a developmental growth continuum (or scale), designed to assist teachers, schools and systems with the process of assessment, recording, reporting and monitoring students' educational progress. For each subject of the curriculum, the framework consists of three key elements:

* areas of knowledge and skills
* aspects within them to be assessed, recorded and reported on, and
* levels of achievement and competency which describe and illustrate both attainment
Figure 1 shows the integration of these elements for one aspect of "Language and Communication".

The framework, as illustrated in Figure 1, reports on the achievements of individual students in an aspect of learning for a given area or subject of learning. Achievements and developing competencies can be shown for the student, a school, a system and nationally. The indicators of competency (whether observations or assessment tasks) have been empirically validated and calibrated on a common scale (using Item Response Theory methodology - Andrich, 1978, 1991; Masters, 1982), thus enabling the use of the full range of assessment methods available to teachers. If indicators are not calibrated in this way, then reliable comparisons by observers on different observations and assessment tasks, or from year to year, cannot be made. Whereas each of the single indicators may refer to a uni-dimensional construct of performance, some indicators and the Profile as a whole reflects the multi-dimensional nature of the skills and knowledge described by that aspect.

The essential feature of a student's Subject Profile is that it shows progression. Through its ordered sequence of achievement levels, the Profile makes explicit what it means to progress in a subject (area) or aspect of learning. It provides a framework against which the progress of an individual can be charted and the achievements of a school or entire education systems can be monitored.

An analogy from Masters (1991) further illuminates the nature of a framework for measuring and reporting on students' achievements in the form of Subject Profiles. He writes:

Measures of temperature are expressed on a temperature scale. This scale provides a framework for the measurement of temperature. The particular numbers used to mark out a framework are a convenience only. Celsius, Fahrenheit and Kelvin scales are common, but other scales are equally possible, and non-numerical approaches to marking out a temperature scale can also be imagined. One way of marking out a temperature continuum is to develop descriptions of observations (eg., behaviours of materials) in particular temperature ranges. When used in conjunction with a numerical scale, these descriptions of typical observations and behaviours give meaning to numbers such as -4°C, 25°C, and 38°C.

Familiarity with the concept of temperature and its measurement enables us to estimate temperatures from everyday observations. These estimates are not always accurate, and nor do they need to be. The estimates we make are often good enough for particular purposes.

To measure temperature more precisely and to monitor small differences and changes in temperature, we use instruments. An instrument (or
assessment task) is simply a way of making more controlled observations of behaviour in an attempt to minimise other, potentially confounding, factors in the observation process. Instruments can be of many different kinds and of varying precisions, but are usually calibrated in terms of a measurement scale.

Finally, once a measurement framework has been constructed, the question of standards can be addressed. In general, a standard is an answer to the question `How much is good enough?' Measurement scales, even when interpreted descriptively, do not define standards. A standard must be defined on grounds external to the measurement process. An example of a standard would be the minimum height for entry into the armed services, or the maximum temperature below which frozen foodstuffs must be stored.

Each of the ideas in this analogy is important in a framework for the measurement of students' current and developing achievements.
Figure 1. An illustrative Subject Profile

First, fundamental to the construction of an assessment and reporting framework is the concept of an achievement continuum, along which individual growth can be measured and charted. For example, positions denoted as 'levels' on the illustrative Subject Profile shown in Figure 1 are indicated using numbers which range from 1 to 10. But the particular choice of numbers is arbitrary and the numerical scale is used as a convenience only.

The example indicators shown in Figure 1 illustrate a second important notion of described levels of attainment or achievement along this continuum. These descriptions and the accompanying assessment tasks play an important role in giving meaning to the reporting scale; moreover, they provide a more informative report on how students are performing and indicate the nature of student growth in writing. Once teachers become familiar with the scale, they can make estimates of the levels at which their students are working on a system-wide and nationally shared framework.

A third important idea is the concept of standard instruments for performance measurement. In brief, valid and reliable performance measures are most likely to result from an assessment system in which instruments are defined more broadly than standardised paper and pencil tests. To maximise educational validity and utility, the instruments used in an assessment program should include commonly-defined activities or tasks which require teacher observations or students to undertake 'authentic' classroom activities, usually with teachers observing and judging individual performance. The Report of the National Curriculum Task Group on Assessment and Testing (TGAT, 1988) in England and Wales explained instruments of this kind in the following terms:

The system that we recommend must include the use of tests, but far more broadly conceived. The aim of the system must be to provide standardised, i.e., nationally comparable, assessment results; but the range and scope of standardised assessments that can be used to perform this role is far wider than the term 'test' is usually taken to imply. We emphasise this fact to draw attention to the fact that our use of the word test will have this broader meaning. Ideally, it might be better expressed by 'standard assessment task'. Both terms will be used interchangeably, to specify externally provided tasks and procedures designed to produce performance data on a national scale.

Instruments are then calibrated against national reporting scales. Instrument calibration (described in more detail later) is a technical process that enables different instruments (sets of assessment tasks, test items or observations) to be used to 'locate' student performance
on the same scale. As noted earlier, if instruments are not calibrated
in this way, then meaningful comparisons on different instruments, or
from year to year, cannot be made.

A fourth important idea is the idea of a standard - an answer to the
question ‘How much is good enough?’ There are different ways of
defining standards and setting goals. One approach is to study a set
of described attainment levels, like those in Figure 1, and to decide
on the level (i.e., kinds of knowledge, skills, and performances) to be
expected of students at a particular age or stage of educational
development. For 17-year-olds, for example, Level 8 from Figure 1 (‘Is
analytical in developing arguments...’) might be considered a minimum
acceptable standard. A goal might then be set to have the majority of
17-year-olds achieving this level by the year 2000.

Finally, the concept of a shared reporting framework is central to
educational assessment programs throughout the world, including recent
Australian statewide performance monitoring programs (eg., Masters et
al., 1990; McGaw et al., 1989; Rowe, 1992a; Titmanus et al., 1993).
For example, the Task Group on Assessment and Testing (TGAT, 1988),
which was established to propose a system of performance measurement
for England and Wales, stressed the need for shared reporting
frameworks which enable individual progress to be tracked from year to
year:

We regard it as one of our priorities to ensure that criteria and
scales used should relate to expected routes of development, giving

continuity to each pupil's assessment at different ages, and thereby
giving a detailed picture of each pupil's progress.... Scales should be
used that indicate where a pupil has reached in a [subject area]. We
shall use the word 'level' to define one of a sequence of points on a
scale to be used in describing the progress of attainment. The
sequence of levels represents the stages of progression.

Having said something about what Subject Profiles `are', `are not' and
perhaps `ought to be', it is both helpful and timely to review the
Victorian experience of the `why' and `how' of Subject Profiles since
1987 when work was begun on the Literacy Profiles project by the
Research and Measurement Centre of the Victorian Ministry of Education.

The Development and Use of Subject Profiles in Victoria

In a paper presented to the Australian Co-operative Assessment Plan
(ACAP) meeting held in Melbourne on 27th June, 1988, Griffin (1988)
provided a detailed account of the historical origins of the
development of Literacy Profiles in Victoria. This account has been
elaborated further by Griffin (1990) and Griffin & Nix (1991). A
summary of the major points from these papers is given here.
From 1984 onwards, Victoria engaged in large scale curriculum reform,
stimulated by several Ministerial papers, and in particular,
Ministerial Paper Number Six (Min. 6), entitled Curriculum Development and Planning in Victoria (1984). Min. 6 urged teachers not to use tests as the sole means of making decisions about student learning outcomes. The paper became the focus of a debate about the use of tests, and the curriculum reform project known as Frameworks moved steadily towards advocating the compulsory abolition of tests in Victorian government schools. Booklets in support of Min. 6 attached an odium to tests which gathered credence among teachers. As a consequence, teachers became confused about whether they were in fact allowed to use any form of testing at all. At least one of the tasks for the Frameworks documents was to change the assessment practices of teachers; but the problem was to change them from what and to what?

In 1986, the Research and Measurement Centre of the then Curriculum Branch surveyed schools to identify teacher assessment practices. The findings from this survey are detailed elsewhere (see Archdall & Kee, 1986; Kee, 1986; Margitta, 1986), but the information gained from this study supported the belief that there was a wide variety of student assessment methods employed across the curriculum and across all year levels. Specific studies of the use of tests (Kee, 1986) and on the reporting practices of primary schools (Archdall & Kee, 1986), identified only isolated instances of inappropriate use of standardised tests, but in the main there was little evidence for the abuse of test information - either at the teacher level or at the school level.

Above all, the findings from these studies suggested the need to employ the assessment skills and methods already used in schools and to highlight these in a way that gave teachers confidence in their own approaches to student assessment.

From the study of reporting practices (Archdall & Kee, 1986), however, the findings suggested a need for the development of better ways for teachers to communicate students' educational progress and attainment information with parents and other stakeholders. Communication was hampered by the range and type of information collected and by the ways this was passed on from teachers to parents and from school to home. Views were sought on parent and employer agencies regarding their expectations of reporting (Murray, 1986; Wood, 1986), and from the literature on student profiling and reporting, particularly from Britain (eg., Broadfoot, 1984, 1985, 1986, 1987).

One paper that aroused interest at the time was by Docking (1986) in which teachers' methods of assessment were cross-referenced against the topics in a secondary school Commercial Studies syllabus. This cross-referencing had several virtues. First, it highlighted the range of assessment practices that classroom teachers were likely to use. Second, it showed that some methods of assessment were more suited to some topics and skills than to others. Third, it showed the potential for describing student competence and performance. Fourth, it provided an organising principle for assembling the variety of assessment techniques and student performance indicators which were pertinent to those techniques.
During this time, the Curriculum Branch was engaged in the literacy and numeracy development programs, supported by the Commonwealth Resource Agreement. This work identified an urgent need for assessment materials to be developed in literacy. While teacher professional development for the Basic Learning in Primary Schools (BLIPS) programs (eg., the Early Literacy Inservice Course - ELIC) were given high priority, there was little being done to help teachers assess the impact of these apparently effective programs. Evaluations of the programs (eg., Rowe, 1987) indicated positive changes in teachers' classroom practices and professional self-perceptions (Rowe & Sykes, 1989), and that teachers firmly believed the impacts were also positive for students. Unfortunately, however, there seemed to be no commonly shared means available to teachers which would assist them to document the alleged changes in students' literacy behaviour. Thus, there was a need for assessment materials that reflected the changes in students' literacy behaviours in such a way that assisted the observation, interpretation, recording and reporting of students' achievements. Recognition of the importance of teacher observation and judgement was considered central to any development Subject Profiles as frameworks for assessing, recording and reporting on students' educational progress. The rationale for this recognition was as follows. (For a more detailed exposition of this rationale, see Griffin, 1989, and Griffin & Jones, 1988).

From the assessment and reporting studies cited above, it was clear that in addition to the use of tests, practical work requirements, assignments, projects, essays, performances, and so on, teachers assess students continuously and intuitively by observation, interaction, questioning, directing, evaluating and supporting students in the process of learning. This formative, analytical and intuitive assessment constitutes one of the most powerful influences on the promotion of students' educational growth and development. Each of these methods of assessment provides a product of an observable behaviour which can be used as an indicator of learning. However, each on its own provides only a small part of the overall picture of student learning and development.

By harnessing and formalising this wealth of information in commonly shared ways of interpreting, recording and communicating the information, it was argued, the relevant audiences (ie., student, parent, teacher, school, system, employer and community) would be better informed about students' established attainments and educational progress. Moreover, the combination of multiple observations from both formal and informal methods of assessment and recording them in the form of Profiles, would provide a more comprehensive view of student progress, and add to the validity of the conclusions reached.

Thus, beginning in 1987, the major aim of the Literacy Profiles project was to make both the formal (tests and related assessment tasks) and the informal (implicit, intuitive and formative) assessment practices of teachers explicit and available to others, so that this powerful information could be systematically gathered, recorded and communicated. It was argued that the major advantage of Subject
Profiles was that they would provide teachers with both a practical and comprehensive domain of competency indicators, using a common nomenclature to facilitate the process of: (1) assessment, (2) recording, and (3) reporting on students' current and developing achievements. While traditional assessment tasks such as tests might be used to inform and assist validation of the assessment process, no one test, or even a battery of tests alone, could be sufficient to meet either the explicit or implicit criteria entailed by the notion of growing competence.

These considerations were influenced by two earlier independent developments in the interpretation of student assessment information. A Danish mathematician, Rasch (1960, 1980), had developed the concept of underlying growth continua or latent traits, which subsequently led to the development of Item Response Theory. Rasch reasoned that the nature of these traits (or attributes) could be defined by the tasks that students performed. If the tasks were to be arranged in order of the increasing amounts of attribute required for successful completion, then the nature of the trait was defined both by the nature of the tasks and their order. Student development could be traced by their progress along the trait or growth continuum. In another development, Glaser (1963) put forward the notion of criterion-referenced testing. This also described performance and development in terms of the tasks performed. Glaser (1981) argued that criterion-referencing should: "...encourage the development of procedures whereby assessments of proficiency could be referred to stages along progressions of increasing competence" (p. 935).

This was to be a vital concept in the development of the Literacy Profiles. It was intended that the project would identify developmental stages of proficiency in reading, writing and spoken language that would enable student performance to be interpreted directly. That is, instead of interpreting the task in terms of individual or group characteristics, criterion-referenced Subject Profiles would allow the interpretation of the individual or group performance in terms of the tasks completed and the behaviours or competencies displayed. The performance or task is thus interpreted by its relative location on a growth continuum and becomes an indicator of development. Thus, Item Response Theory can be used as an empirical means of testing the fit of the criterion-referenced tasks and their placement order along the developmental growth continuum.

Against this background, eight principles for recording and reporting student assessments in the form of Subject Profiles were proposed (see also Griffin & Nix, 1991:2). It was argued that such Profiles should be:
1. Valid and reliable descriptions of student learning outcomes
2. Criterion-referenced predominantly and based on a defined construct
3. Progressive and developmental, reflecting an underlying growth continuum
4. Analytical in that both strengths and weaknesses are identified
5. Based on direct observations of behaviour
6. Flexible in that it does not control the curriculum or teaching methods
7. A communication of attainments and developing achievements, with realisable expectations of progress
8. Moderated through teacher meetings or conferences.

Hence, several factors coincided. First, there was a need to demonstrate that teachers used a variety of assessment methods and to encourage them to identify the appropriate circumstances for these different methods, including tests. Second, there was a need for assessment techniques and instruments to be described and developed, first in the area of early literacy (Prep to Grade Three) and then later for Years 4 to 10, that reflected students' educational growth and developing competence. Third, there was a conviction that cross-referencing assessment techniques with types of learning was a useful organising principle. Fourth, there were several factors identified in the assessment and reporting studies that needed to be put into practice. Fifth, there was a need for indicators of achievement to be expressed in commonly shared and understood nomenclature that could be used for teacher, school, parent and system-level purposes.

These factors gave rise to the construct of a matrix, designed to cross-reference the method of gathering assessment information with the stage of student development outcomes for a given area of learning - what teachers subsequently referred to as "milestones". But instead of merely associating these two sets of important information as Docking (1986) had done, the specific performance indicators in the cells of the matrix became the focal point. The assessment methods were to be matched to each outcome so that each cell of the matrix would contain a description of behaviour (indicator) which could be observed and verified as indicative evidence of both development and attainment. These were to be combined in a simple worksheet as shown in Figure 2. The procedure used in identifying the 'milestones' for the Literacy Profiles (Victoria, 1990) (now known as the English Profiles - Victoria, 1991), their performance indicators and the subsequent bands of development is instructive, as it may be applied to a variety of areas in which performance indicators are needed. The procedure also illustrates the connection between indicators and outcomes.

Figure 2. Matrix showing performance indicators aligned with expected outcomes and related assessment methods
The Procedure Used in Developing Subject Profiles

1. Teacher Workshops. For the initial development of the Victorian Literacy Profiles during 1987, approximately 100 primary teachers, divided into four groups, spent four days spread over the year working in syndicates of six, in structured workshops. (For a more detailed account of these workshops, see Griffin, 1990, Griffin & Jones, 1988, and Griffin & Nix, 1991). The project was initially restricted to primary classroom teachers because of the Ministry of Education's priority focus (at that time) on the assessment of literacy development at the junior primary levels. The workshops used an analytical method which combined the identification of areas of reading and writing development, the delineation of appropriate outcomes associated with each area, and a range of assessment methods which could be used to provide evidence that the outcomes had been achieved. The evidence obtained by each assessment method, called a "performance indicator", was written into the relevant cell of the matrix worksheet. Figure 3 below shows how these "performance indicators" were combined into a worksheet.

Twenty four matrix worksheets were developed covering a range of literacy areas. The example shown in Figure 3 illustrates the relationship between the area of developing an "Approach to Unknown Words", the outcomes and the relevant assessment methods. (For a complete version of this matrix, see Griffin, 1990:292).

<table>
<thead>
<tr>
<th>Assessment Techniques</th>
<th>OUTCOMES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asks others</td>
</tr>
<tr>
<td>Observation and already anecdotal stories or oral</td>
<td>Asks adult/peers</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation and already anecdotal stories or oral</td>
<td>Asks adult/peers</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Running record</td>
<td>Asks adult/peers for meaning &amp; pronunciation of words</td>
</tr>
<tr>
<td>Parent/teacher</td>
<td>Asks parent/teacher for States that the picture States a</td>
</tr>
<tr>
<td>Parent/teacher</td>
<td>Asks parent/teacher for States that the picture</td>
</tr>
<tr>
<td>Parent/teacher</td>
<td>Asks parent/teacher for States that the picture</td>
</tr>
<tr>
<td>Parent/teacher</td>
<td>Asks parent/teacher for States that the picture</td>
</tr>
</tbody>
</table>
The word is "known" conference meaning and pronunciation helped in reading text because it "sounds right"

Figure 3. Matrix worksheet: "Approach to Unknown Words"

2. Observation and Moderation. Back in their own or other classrooms, teachers collected and/or documented examples of the students' reading behaviours that were then panelled and discussed at a subsequent workshop session. Each group assessed the exemplar materials until consensus was reached whether these represented appropriate evidence for the outcomes. Where agreement could not be reached, the indicator was removed from the matrix worksheet. Only those indicators that were agreed upon by groups of teachers were retained in the worksheets. Moreover, only those statements of indicators which could be observed directly were retained. The process of describe, observe, discuss and revise was an integral part of the development of initial indicator statements. A workshop facilitator visited every teacher who had participated in the project and had assisted in completion of the matrix worksheets and moderation of work samples for indicators. These visits helped to achieve a degree of standardisation among the observational data being collected; and moderation among peer groups of teachers during workshops proved to be productive in developing statements of indicators. Comparisons of work samples, anecdotal records and other formal and informal assessment results assisted teachers in agreeing upon the final form of articulation for the indicator and the development of the matrix worksheet.

3. Scale Development. A set of 24 matrices describing reading behaviours were used by teachers to record their observations of children. Observations were made in a survey conducted in 22 schools, involving a total of 332 students in Grades 1 to 5. At the instigation of teachers, a rating scale was introduced to indicate the level of confidence they had that the indicator behaviour was exhibited. Teachers correctly argued that the development of these behaviours was more complex than a simple "present/absent" rating of a student's reading behaviour. There had to be a way of accounting for circumstances in which a performance indicator had yet to become established for a student, but was nonetheless observable on some occasions. By introducing a rating scale using 0, 1 and 2 to describe increasing evidence of an indicator being exhibited, teachers allowed for this possibility. Thus, for each indicator a zero (0) rating was used if the teacher had not observed a student exhibiting a performance indicator; a rating of one (1) was used to indicate that the behaviour was still developing and had yet to be consistently displayed by the student; and a two (2) was used if the teacher was confident that the
indicator was an established part of the student's repertoire of reading behaviours. Data from the survey were analysed using Item Response Theory (IRT) technology (Rasch, 1960) as it applies to partial credit ratings (Andrich, 1978, 1991; Masters, 1982). This enabled the indicators to be calibrated so that they could be mapped onto a developmental scale or growth continuum. This scale is expressed in terms of analytical indices known as logits, which typically range in value (in order of item difficulty and in order of increasing developmental attribute) from -5 (easiest and lowest level of attribute development) to +5 (most difficult and highest level of attribute development), as illustrated in Figure 4.

Figure 4. A progression indicators mapped onto the developmental scale or growth continuum

Such IRT approaches to item analysis assume that the attribute of reading competence (in this case) is an attribute possessed by the
student and that there are reading-related tasks which require
different amounts of this attribute in order to become established
parts of a student's behavioural repertoire. That is, each reading
task has a required attribute level and can be located on the
developmental scale of increasing attribute. If the student possesses
more of the attribute than a given task requires, it is more likely
than not that the student will be able to perform the task, and that
the reading-related behaviour described by the indicator will be
established in the student's reading behaviour pattern. If the task
(or its performance indicator) requires more of the attribute than the
student has so far been able to develop, it is more likely that the
indicator will not be established part of the student's reading
behaviour.

Calibration of the indicators yielded some interesting results. All of
the indicators which described reading behaviours outside of teachers'
direct observations were found to have erratic rating patterns. Such
indicators were those which described the students' behaviour at home,
borrowings from public libraries and other reading activities outside
the classroom. These were excluded from the developmental scale. The
remaining indicators were ordered empirically according to the amount
of attribute required (logit value) as shown in Figure 4, which
represents a selection of the indicators from the scale range.

It is important to note that the examples shown in Figure 4 do not
represent an exhaustive list of behaviours to observe in assessing
reading development. The same is also true for the more complete
versions that were first published in the Literacy Profiles Handbook
(Victoria, 1990) and later in the English Profiles Handbook (Victoria,
1991). Rather, they constitute a selection of indicators from the
calibrated list, spread over the full range. The danger of presenting
the indicators in checklist form such as this, however, is that
"washback effects" may arise. That is, the checklist can become the
"rule" for teaching as well as the "ruler" for monitoring. Griffin
(1990:296) suggests that it may be better for teachers to develop their
own checklists which match their curriculum and are appropriate to
their specific needs. Otherwise, the sequence of indicators can be
erroneously interpreted as the sequence for learning and even teaching,
resulting in a "wash-back effect" from assessment to teaching.

Producing a long checklist, moreover, helps neither the monitoring nor
the analysis processes. It is clear that the indicators are spaced
very closely in terms of assessing development and hence it was not
necessary to use them all. Furthermore, setting hard and fast rules
about which indicator precedes which can be challenged on a large
number of legitimate grounds. Similarly, because of the context,
content and purpose dependency of reading behaviour, the relative
positions of the indicators could be unstable.

4. The Determination of Bands. The full list of indicators obtained
from the analysis of the matrices was examined in terms of their
respective logit values on the developmental scale, and for patterns which might be useful in summarising them into groups which were subsequently labelled as 'bands' or 'levels'. A similar exercise had been undertaken in the aggregation of indicators for the Australian Second Language Proficiency scales (Ingram, 1984). This examination yielded several clusters. For example, the set of eight indicators shown in Figure 4 appeared to be a logical progression. They all fit the growth continuum and seem to describe increasing competence to analyse and identify elements of textual information.

Several progressions were identified in the list of calibrated indicators, and all of these progressions could be shown to fit empirically into an overall developmental growth continuum of reading behaviours. The task was to put them into a useable form. The progressions seemed to be related to underlying factors such as attitudes towards reading, influence of reading on writing, role playing, retelling, reactions to reading materials, analysis and interpretation, social or interactive roles in reading behaviour, word attack skills, types of reading material used, and so on. However, not all of the progressions could be traced all the way through the list. Some were restricted to the lower end of the developmental scale, some emerged only at the upper end, and some appeared to be absent at the extremes.

The groups of indicators were called "Bands", each of which contained a broad description of a range of reading behaviours rather than providing a discrete definition of specific points of development. Seven Bands were developed initially and were later extended to nine. (For specific details of the indicators grouped into the Bands, see Victoria, 1990, 1991; an example is also given in Appendix 1). It should be noted that a Band does not represent anything other than a grouping of indicators with a common logit value range on the developmental scale. Moreover, determination of the number of Bands is purely arbitrary. For example, in the interest of providing finer discriminations in order to show student progress throughout a school year, or from one year to the next, there may be compelling reasons for increasing the number of Bands. Nevertheless, it should be stressed that the Bands certainly do not represent expected levels of competence for designated years of schooling.

The Bands were labelled `A' to `I', with each group of indicators having a value range of approximately 1.5 logits. Band A was deliberately set at the earliest developmental level in order to avoid the association of "value" with development, and to enable the scale to be extended later with further Band levels for use by teachers in secondary schools. Due to the empirical calibration of their constituent indicators using IRT technology, the Bands were designed to be cumulative. This is an important feature of the English Profiles which indicates that for a student whose mother tongue is English and is developing competencies at Band E, the student would also be likely to have established the behaviours described at lower band levels, but
unlikely to have established competency at higher levels. For non-English speaking background students with well-developed reading skills in their mother tongue, trials of the Reading Profile Bands indicated a pattern which showed "beginning" development across several band levels (see Rowe, 1991a).

Draft forms of the Reading Profile Bands (and later the Writing and Spoken Language Profile Bands) were taken to schools for discussion with teachers, followed by extensive consultation with experts in literacy education for refinements of expression and presentation format. Identical processes to those described above were adopted for the identification, calibration, moderation and refinement of the indicators of written and spoken language.

5. Field Trials and Research. In 1988 the opportunity arose to formally trial the Reading Bands as a monitoring and reporting instrument in the 100 Schools Project - Literacy Programs Study (see Rowe 1990, 1991a, 1991b, 1991c). In this three-year longitudinal study of factors affecting students' achievements in reading conducted between 1988 and 1990, the Reading Bands were used by classroom teachers on more than 5000 students drawn from 100 government, Catholic and independent primary schools and secondary schools. An important feature of this study was the use of domain-referenced tests of reading to assist teachers in recording students' progress on the Reading Profiles.

During 1991 both the Reading and Writing Bands were trialed in the Victorian Profiles Program among a sample of 5100 Year 6 students from 134 government, Catholic and independent schools (see Rowe, 1992a, 1992b). In 1992 Victoria again used the English Profiles (reading, writing and spoken language) together with the Mathematics Profiles (space and number) to monitor the achievements of a statewide sample of 8500 Year 3 and Year 6 students in 239 government and Catholic schools (see Victoria, 1993). However, the most extensive use of Subject Profiles to date has been in the Victorian Quality Schools Project (Hill, Holmes-Smith & Rowe, 1993a, 1993b; Rowe, Hill & Holmes-Smith, 1993, 1994). From this three-year longitudinal study of teacher and school effectiveness (1992-1994), data on educational progress in English and Mathematics have been obtained for entire year level cohorts (P-10), made up of a sample of 13,900 students drawn from 90 government Catholic and independent primary and secondary schools.

Further, beginning in April 1994, the Victorian English and Mathematics Profiles are being used as part of a four-year longitudinal study to assess the impact of the Victorian government's Schools of the Future policy (Victoria, 1994) on aspects of school effectiveness, involving 10,400 students in 90 government primary and secondary schools. For these projects, it should be noted that following specific advice from classroom teachers for the purposes of recording and monitoring student progress on the English and Mathematics Profiles, a further rating scale has been developed - modelled on those described in the
book Reading On (Victoria, 1984). The levels of development in that text were identified as beginning, developing and established. Teachers' familiarity with these terms and with the notions of global, holistic and impression monitoring, meant that familiar constructs have been able to be used for rating each Band or Level. This has resulted in a four-point scale as follows: 0 - no evidence, 1 - beginning, 2 - developing, and 3 - established. Examples of the Recording Instructions for the English and Mathematics Profiles used in the Victorian Quality Schools Project (Hill et al., 1993a) are given in Appendix 2.

How Subject Profiles Can be Used

On the basis of evidence from a variety of "in-context" formal and informal methods of assessment, teachers in the studies cited have typically recorded students' progress by constructing class lists of the kind shown in Figure 5, using the method for rating the descriptive indicators of the Bands (or Levels) mentioned above.

Figure 5. A partial class list of student progress in literacy, from the Victorian Quality Schools Project

For the purpose of reporting to parents on occasions such as parent-teacher interviews, these 'ratings' can be presented graphically in a form as shown in Figure 6. A collection of such forms over term periods, semesters, or over several years, may also be included in a student's "Record of Achievement". Feedback from teachers and schools who use such graphical forms in reporting students' educational progress to parents continues to be overwhelmingly positive.

Figure 6. Example of an individual student's progress on a Reading Profile

A particular advantage in using a criterion/domain-referenced reporting method of this kind is that the 'nutshell' statements (enclosed in the boxes of Figure 6) act as 'pointers' to the more detailed descriptive indicators which they represent (see Appendix 1). Moreover, parents (and other interested stakeholders) can be shown actual samples of work, including the results of class-based assignments, tests and other assessment tasks that illustrate students' current and developing 'locations' on the Profile scale. Mention was made earlier that Subject Profiles can also function effectively as standards monitoring frameworks at the school, system and national levels, provided they are "established empirically" (Hill,
1994:38). This has been possible in the case of the Victorian English and Mathematics Profiles which have been used extensively in large-scale monitoring and explanatory research studies (see citations mentioned on pp. 15-16). To illustrate this use of the Profiles, Hill (1994) presented data for reading achievement levels of students in each of Years Prep, 2, 3, 4, 6, 7 and 9 on the reading strand of the English Profiles (Victoria, 1991). A further illustration is provided here using data for students rated on the number strand of the Mathematics Profiles (Victoria, 1992).

Whereas the English Profiles are organised into nine levels or `bands' (labelled `A' to `I'), the Mathematics Profiles consist of 12 levels labelled `1' to `12'. Teacher ratings for each student on the 12 levels (using the codes 0, 1, 2 and 3, as outlined above and described in greater detail in Appendix 2) have been added together to give a total `score' out of 36. Figure 7 shows the achievement levels of students in each of the Years Prep to Year 10, in the form of 'box and whisker' plots (Tukey, 1977). Similar data for students' levels of achievement on each of the strands for the Victorian English and Mathematics Profiles are given in Appendix 3.

Figure 7. A standards framework for number

The shaded boxes in Figure 7 describe the range of achievement of the `middle' 50 per cent of students at those Year levels. The top of each box indicates the level of students achieving at the 75th percentile, the bottom of the box shows the 25th percentile and the asterisk indicates the 50th percentile. The top and bottom `whiskers' show the 90th and 10th percentile levels of achievement respectively.

The data summarised in Figure 7 indicate a steady and consistent rate of growth in students' achievements in number from Prep. to Year 10. It is noticeable, however, that there is increasing differentiation in the range of achievement from primary to secondary schooling, indicating that secondary teachers are required to manage teaching and learning processes for a wider range of student competence than that required of their primary colleagues. As discussed in more detail by Hill (1994:38-39), this phenomenon is particularly marked in the comparable data on the reading, writing and spoken language strands of the English Profiles, as well as on the space strand of the Mathematics Profiles (see Appendix 3).

Information of the type shown in Figure 7 (and those given in Appendix 3) is of interest in its own right by providing a `picture' of student progress within given curriculum areas and strands. As noted by Hill (1984:39): "It is also essential to enable the profiles to be used to report student achievement and progress in terms of standards typical of students of the same age or year cohort." However, this raises several issues related to the potential use of profiles.
Potential Uses of Subject Profiles

In the experience of the present authors, Subject Profiles have been used most profitably in both monitoring studies (Rowe, 1992a, 1992b, 1994; Victoria, 1993) and in longitudinal explanatory research projects (Hill et al., 1993a, 1993b; Rowe, 1990, 1991a, 1991b; Rowe et al., 1993, 1994; Rowe & Rowe, 1992a, 1992b). In each of these studies, information obtained from the profiles have been used at the levels of teacher, parent, school and system. At the teacher level, such information can be used diagnose student strengths and weaknesses, and to identify teaching and learning areas requiring further attention (see Figure 5). This is accomplished by examining individual student and/or class development on specific indicators in the 'bands' or 'levels' and setting targets for student learning. The particular utility of the developmental growth continuum across the years of schooling (see Figure 7 and Appendix 3) is that it provides teachers with a 'map' which facilitates the specification of individual student and/or group learning goals. In fact, as noted by Hill (1994), there is considerable interest among teachers in exploring the use of profiles as a means of improving teaching and learning processes in the classroom. Several other ways in which teachers can make use of profiles is described by Wilson (1993).

As mentioned earlier, an important use of profiles is in reporting to parents. One method of doing this is illustrated in Figure 6. While this ipsative and criterion/domain-referenced approach has been well received, parents frequently require additional norm-referenced information of the kind expressed in the common query, 'How is my child going relative to standards for this Year level?' (Archdall & Kee, 1986). By referring simultaneously to data of the kind presented in Figures 6 and 7, teachers can respond to such queries with confidence. Hill (1994:40) stresses the importance of this use of profile information in the following terms:

...profiles allow teachers to communicate to parents about student progress and achievement using a language and standards which are consistent across classrooms, schools and school systems, and, where empirical data are available, in relation to performance levels typical of other students of the same age or year level.

Such norm-referenced information can also be used to monitor the performance levels of both individual students and Year level groups, to ensure that progress is being made and that performance levels are within expected standards for relevant year or age cohorts. At the school level, by compiling profile results over several years, it is possible for schools to determine trends in standards and performance levels over time. To illustrate this, Hill (1994:41) cites an example from the Victorian Quality Schools Project (Hill et al., 1993a):
...aggregated school-level assessments obtained using teacher ratings of student achievement based on the Victorian profiles, have been provided to 90 participating schools in a form which enables them to compare their achievement levels with State standards and also with standards of schools with a similar student intake as measured by family socio-educational level. This information has been used by the schools in the context of introducing quality assurance practices aimed at improving educational outcomes (see Rowe et al., 1993).

It is also possible to make use of profiles for monitoring educational standards at system and national levels. Several such initiatives are currently in progress, including the Western Australian Monitoring Standards in Education Program (Titmanus et al., 1993; Western Australia, 1991). A feature of this program is the simultaneous use of standards frameworks in the form of Subject Profiles for English and Mathematics, linked empirically to specially constructed tests and assessment tasks using Item Response methodology. Student results on these tests and assessment tasks are reported in terms of their 'locations' on the achievement levels indicated on the growth continuum of the standards framework. This kind of approach is particularly appropriate in instances of high stakes, system-level monitoring of standards, where it would be unwise to rely entirely upon school-based teacher assessments using Subject Profiles, since these may be artificially inflated or reflect possible inconsistencies in the interpretation of standards across teachers and schools. In such situations, the use of context- and domain-specific performance testing and related assessment tasks of the kind currently being developed by the ACER are required (see Fortster, 1994), to facilitate the reporting of student achievement in terms of the levels delineated by Subject Profiles.

In low stakes explanatory research studies, however, the use of school-based assessment methods linked to Subject Profiles have been shown to be a highly cost-effective means of obtaining comparable data on standards and for estimating the effects of factors influencing students' educational progress. Moreover, from our experience, this use of Subject Profiles has had positive impacts on schools in terms of encouraging them to respond to the challenges of evaluating programs, monitoring student progress, and adopting specific strategies aimed at school improvement (see Rowe et al., 1993).

Concluding Comments

There are three important points to stress from this work on the development of Subject Profiles and their uses as standards frameworks for assessing, recording and reporting on students' educational progress in both monitoring and explanatory research studies:

1. First, the distinctive feature of Subject Profiles is that what is to...
be achieved is stated explicitly. In the constituent aspects of each subject area, for example, the nature of 'writing' (Figure 1), and 'reading' (Figure 6), is specified, as are the levels or benchmarks which students progressively attain on their path to proficiency. The full range of achievement from early stages to mastery is shown, regardless of how that achievement has been assessed.

Second, higher order outcomes of knowledge and skills are specified. There is an attempt to go beneath the surface variations of the curriculum delivered in classrooms and state those things which are important for all students to know and be able to do.

Third, and perhaps above all, Subject Profiles allow the full range of both formal methods of student assessment (tests and related assessment tasks) and informal methods (observations and descriptive judgements) typically used by teachers, to be calibrated and mapped onto a common developmental scale or growth continuum. Using recent developments in Item Response Theory methodology (eg., Adams & Khoo, 1992; Andrich, 1991; Masters, 1988), the major advantage of such scales is that they enable students' performances on different tasks and in different subjects to be compared reliably - from student to student, from year to year, and over time.

References


Committee on Assessment (1990). Curriculum and Assessment in Scotland:


Hill, P.W., Brown, T., & Masters, G.N. (1993). Fair and authentic school assessment: Advice to the Board of Studies on verification, scaling and reporting of results within the VCE. Melbourne, Vic: Board of Secondary Education.


O'Connor, M.C. (1992). Rethinking aptitude, achievement and


Rowe, K. J. (1991a). The influence of reading activity at home on students' attitudes towards reading, classroom attentiveness and reading achievement: An application of structural equation modelling.

British Journal of Educational Psychology, 61 (1), 19-35.


Appendix 1: The Reading Bands (from Victoria, 1991, pp 86-87)

Appendix 1 (Cont.)

Appendix 2.

ENGLISH PROFILES
READING, WRITING & SPOKEN LANGUAGE BANDS:

RECORDING INSTRUCTIONS
Read the descriptions of behaviours in PROFILE BANDS labelled A to I. Compare each student's "Reading", "Writing" and "Spoken Language" behaviour patterns to those described for each BAND (in "The Bands at a Glance"), and use the following codes:

3-If the student has established the behaviour pattern, and consistently exhibits all or most of the behaviours in the BAND, use a code of 3.

2-If the student is developing the behaviour pattern, such that some, but not all of the behaviours for a band are often exhibited, use a code of 2 for that BAND.

1-If the student is beginning to show signs of the behaviour pattern for a BAND level, in that only a little of the pattern is shown, use a code of 1 for that BAND.

0-If the student shows none of the behaviour pattern for a BAND level at this stage, use a code of 0 for that BAND.

EXAMPLE

Using the coding method outlined above, enter a code in each BAND in the boxes provided on the STUDENT ACHIEVEMENT & BEHAVIOUR RECORD form. The Profiles for one student are given below as an example:

INTERPRETATION of Profiles

For Reading, the student has established the behaviours described in BANDS A to E, is developing the patterns in BAND F, and is beginning to show signs of BAND G behaviours. The student has not yet begun to show the behaviours described in BANDS H and I.

For Writing, the behaviours described in BANDS A to C have been established and the patterns in BAND D are developing. The student is beginning to show signs of BAND E and F behaviours, but has not yet begun to show the behaviours described in BANDS H and I.

For Spoken Language, the student consistently displays the behaviours described in BANDS A to C, is developing the behaviours described in BAND D, and is beginning to show signs of BAND E behaviours. There is no evidence of behaviours described in BANDS F to I.

Centre for Applied Educational Research
The University of Melbourne
Appendix 2 (Cont.)

MATHEMATICS PROFILES
NUMBER AND SPACE LEVELS:

RECORDING INSTRUCTIONS

Read the descriptions of behaviours in PROFILE LEVELS labelled 1 to 12. Compare each student's "Space" and "Number" behaviour patterns to those described for each LEVEL (in "The Levels at a Glance"), and use the following codes:

3-If the student has established the behaviour pattern, and consistently exhibits all or most of the behaviours in the LEVEL, use a code of 3.

2-If the student is developing the behaviour pattern, such that some, but not all of the behaviours for a level are often exhibited, use a code of 2 for that LEVEL.

1-If the student is beginning to show signs of the behaviour pattern of a LEVEL, in that only a little of the pattern is shown, use a code of 1 for that LEVEL.

0-If the student shows none of the behaviour pattern for a LEVEL, use a code of 0 for that LEVEL.

EXAMPLE

Using the coding method outlined above, enter a code in each LEVEL in the boxes provided on the STUDENT ACHIEVEMENT & BEHAVIOUR RECORD form. The Profiles for one student are given below as an example:

INTERPRETATION of Profiles

For Space, the behaviours described in LEVELS 1 to 3 have been established and the patterns in LEVELS 4 and 5 are developing. The student is beginning to show signs of LEVEL 6 behaviours, but has not yet begun to show the behaviours described in LEVELS 7 to 12.

For Number, the student has established the behaviours described in LEVELS 1 to 4, is developing the skills in LEVEL 5, and is beginning to
show signs of LEVEL 6 to 8 behaviours. The student has not yet begun to show the behaviours described in LEVELS 9 to 12.

_______________________
Centre for Applied Educational Research
The University of Melbourne

Appendix 2 (Cont.)

Appendix 3.

Appendix 3 (Cont.)

Appendix 3 (Cont.)

Appendix 3 (Cont.)

Appendix 3 (Cont.)

Appendix 4. Reliability Properties of the Victorian English and Mathematics Profile Strands

Using the holistic rating method for each ‘band’ or ‘level’ described on page 16 and shown in Figure 5 (see also Appendix 2), a rating pattern for each student forms a qualitative cumulative scale similar to that described by Guttman (1944). Reliability analyses have been computed for each strand, by Year level, using the Guttman method to provide lower bound estimates of true reliability. The relevant coefficients given in the Table below are standardized item alpha
estimates and refer to data obtained from the Victorian Quality Schools Project (see Hill, Holmes-Smith & Rowe, 1993 a). The sample sizes (n's) for each Year Level cohort are given in parentheses.


See Broadfoot et al. (1990), OECD (1993).

Examples of this expertise are illustrated by Griffin (1990), Griffin & Nix (1991); by the Victorian Ministry of Education's Profile Handbooks (Victoria, 1990, 1991, 1992); by the Australian Council for Educational Research in the Victorian Achievement Studies (Adams et al., 1991; McGaw et al., 1989) and in the Developmental Assessment Resource for Teachers (DART - Forster, 1994); the New South Wales Basic Skills Testing Program (Masters et al., 1990); the Western Australian Monitoring Standards in Education Program (Titmanus et al., 1993) and Student Outcome Statements (Western Australia, 1993), the Queensland Assessment of Student Performance (1990); the National Profiles projects (see Hill, 1994); the National Statement on Mathematics for Australian Schools.