

Monitoring student learning using assessment data.

Gina Silis and John Izard

Monbulk Primary School and RMIT University

SIL02449

Presented at the AARE Conference in Brisbane

December 2002

1. Context of the study

Monitoring of student learning often refers to activities of agencies external to schools including comparisons of schools or comparisons of nations. The study reported in this paper is different. This study involved those who delivered the learning and they did the monitoring of the learning. They obtained evidence of each student's present level of achievement by assessing all students periodically. Results were not used to compare students (or teachers). Results were used to identify what each student knows so that future learning could be planned. This future learning was believed to be more effective because it was based on what an individual currently knows and targets skills that were within reach of that student.

In this section we describe the situation that triggered the development of an innovative program initiated by a school in 1996. The program developed from a need for achievement data to be useful to classroom teachers and supportive of their primary aim to improve student learning. This was in contrast to the action and approach to the monitoring of student progress by Government outlined below.

In the early 1990's Government Primary Schools became part of a State and National monitoring of achievement program. In the State of Victoria it is termed the Achievement Improvement Monitor (AIM). AIM is comprised of the following components. Classroom Assessment, Comprehensive Reporting, Homework Guidelines and Learning Improvement.

Standardised tests are used by the State to monitor learning improvement. The tests are administered to students in Grade 3 and 5 in Primary Schools. Testing occurs in the middle of the school year and is carried out by a student's teacher. Tests are then collected and marked centrally. Results are returned to schools in the closing term of the school year. The AIM test covers the curriculum areas of Mathematics and English.

The AIM is linked into the National Goals for Schooling, a document finalised in 1999 by the Ministerial Council for Education, Employment, Training and Youth Affairs (MCEETYA). The impetus arose from "Goal 2.2 In terms of curriculum, students should have: attained the skills of numeracy and English literacy; such that, every student should be numerate, able to read, write, spell and communicate at an appropriate level". (MCEETYA (b), p. 4) A related sub-goal stated: "every child commencing school from 1998 will achieve a minimum acceptable literacy and numeracy standard within four years". (MCEETYA (a), p.11) The above MCEETYA statements imply the collection of assessment data to provide the required evidence.

In 2000, the Australian Department of Education Training and Youth Affairs (DETYA) produced *Numeracy, a priority for all: Challenges for Australian Schools, Commonwealth Numeracy Policies for Australian Schools*. This document established the need to develop benchmarks outlining minimum achievement expectations and the location of them within the State tests. "School authorities are working through an agreed national process to locate the benchmark on the tests, to enable nationally comparable reporting of aggregated performance data by States and territories." (DETYA, p.22)

1. Purpose of assessment

The purpose of assessment is described by Lai and Griffin (2001) as engagement in procedures that aim to gather and interpret evidence regarding achievement. Making valid judgements regarding the learning that has taken place is pivotal. William (1996) states that "validity is therefore not a property of tests, nor even of test outcomes, but a property of the inferences made on the basis of these outcomes."

There are at least three forms of assessment models (examinations, state and national programs, teaching and learning). They each have a specific purpose. The high stakes *examinations* contain many items and often a mixture of multiple choice, short answer or essay responses. They are described as high stakes because of the way in which the results are used (for example, University entry, job placement and use of the ranking scores). *State and national testing* are also high stakes programs. They are carried out mid-year to late in the school year and provide cohort information. The results are used for monitoring and accountability purposes. These programs are described as high stakes because of the sanctions that exist for schools perceived as ineffective. *Teaching and learning* programs are concerned with identification of the learning needs of individuals in a grade and need to support planning and instruction. Those lacking knowledge of the learning process do not consider them as high stakes programs. But teaching and learning are essential elements that contribute to success on high stakes programs. We believe there should be greater emphasis on the use of assessment for teaching and learning. The focus of this paper is on the latter.

2. Concern regarding State wide testing program and accountabilities.

The shift in focus to accountability, and standards monitoring, has detracted from the potential to provide meaningful and valid information to teachers, parents and administrators regarding student-learning requirements. There is a possibility that school personnel are unable to translate these new statistical images of performance validly into improved learning for students. In our view, years of focus on informal assessment have left a void in understanding and skill among teaching staff regarding the applicability of formal objective assessment tools. To a considerable extent, these skills are held by a diminishing group of aging teachers. They include those trained to handle specific learning difficulties as a result of a "Federal House of Representatives Standing committee on Specific learning Difficulties, set up in October 1974, at the absence of reliable evidence on the extent and distribution of specific learning difficulties in the Australian population." (Piper, 1997, p.97)

Teachers in the 1980's had been encouraged to move from informal testing strategies to subjective assessment strategies, for example annotated notes,

checklists, projects, journals, portfolios, investigations and open-ended problem solving tasks checklists. Teachers were provided with professional development in how to construct and develop these strategies. The professional development opportunities explained the problems that may arise from relying on these forms of assessment as primary methods of collecting information relating to student achievement. Unless the professional development continues, unfounded assumptions and failure to consider alternative explanations lead to suspect conclusions.

A statewide testing program is applied from above and this causes concern for many teachers and school administrators. Very few, if any, teachers have been involved in the development of the tests or procedures. The State tests administered in the middle of the school year are not seen by teachers as useful as the results do not arrive back in schools until close to the end of the school year. Piper (1997) describes the assessment of achievement in Year 3 and Year 5 of a students' primary educational experience as spasmodic. "Further, the assessment is not in terms of which curriculum objectives have been achieved. The assessment is in terms of status of a particular group of students relative to the proportion of reference group whose qualities are not known to the teachers." (Silis & Izard, 2001, p. 5)

For this reason, few schools have utilised externally developed and published tests to monitor student achievement and support *teaching and learning*.

3. Terminology

With the advent of the State and National Monitoring of student progress, statistical terms, such as percentile ranking (sometimes reported as expressed as "box and whisker" graphs), cohort, median, mean and standard deviation needed to be comprehended by teachers. The meanings of these terms and the use of the concepts to provide an accurate interpretation of scores are further challenges for teachers to comprehend and master.

Wright and Stone (1979) expressed concern that "percentile ranks and standard scores" (p. viii) are not adequate measurements of performance on a test paper or between test papers of ability. Group difference can be noted but not individuals. A scale of correlation needed development that involved the difficulty of the actual questions (items) on the test, relating items and students. Griffin (1998) states, "Equating is a process that brings different measures of the same construct into alignment so that all the items in an item pool, and the students who attempt them, can be described with the same units of measurement" (p.4). Griffin goes on to state that the procedures used in Item Response Modelling (for example, Rasch Models) enable equating methods that establish equivalence of tests and sub-tests used with a range of groups. These equating methods are preferred to classical test model because they;

i give item statistics that are not group dependent:

ii yield scores describing examinee proficiency that are independent of test difficulty, and

iii do not require strictly parallel tests for establishing equivalence.

(Lord, 1980; Wright & Stone, 1979; Black & Wiliam, 1998a;
Black & Wiliam, 1998b.)

Non-statistical terms such as 'Like' school are also used. Schools are grouped for comparison on two criteria. The proportion of students for whom the main language spoken at home is not English, and the proportion of students for whom parents receive the educational maintenance allowance (EMA). (VCAA, 2000, p8)

In 2001 Virgona and Greaves pointed out "Educators have a role in encouraging class teachers who are uncomfortable with no more than total score to consider the real value of Item Response Modelling tests." (p. 42) Further, Black and Wiliam (1998) have pointed out that improving student learning is a key teacher role and therefore formative assessment and consequent action by the teacher and students is essential.

4. Forms of assessment

Subjective assessment occurs in formal and informal ways in classrooms using a range of strategies ranging from annotated notes, checklists, projects, journals, portfolios and investigations to open-ended problem solving tasks. Included here are teacher-designed tests. In designing one's own measurement tools, teachers tend to reflect individual strengths and weaknesses and the curriculum they believe has been covered. This makes it difficult to use such information for directing required teaching. For example, the ability to develop several multiplication items to cover the range of progressive steps involved in this operation requires time and skill and trial of the items. Do the items show the teacher how much or how little all the students in the grade know? Have all the technicalities of the process been covered?

Moderation of teacher judgements is a large issue in schools. When assessment is subjective it takes teachers a great deal of time to develop consistency between judgements. A balance of assessment forms is needed to verify teacher judgement and provoke discussion. For assessment to be informative the information must be understood and used.

5. Comparative assessments.

The assessing of progress in learning implies that a series of measurements will be taken to provide evidence of the changes that are expected to occur. The tools used to take these measurements need to relate to each other. They need to measure the same construct (Mathematics, or English -Reading and Spelling) and cater for range of abilities.

Schools have a need to understand and interpret the data that they are asked to collect and convert into information that results in improved outcomes for students. The curriculum is a statement of intention and unless we gather data we have no evidence that the teaching program is delivering the necessary learning experiences. Evidence of delivery and change require at least two occasions to be measured. In the 60's Item Response Theory was nothing more than a theoretical possibility from the perspective of a classroom teacher. The potential it offered was recognised. Lord and Stocking (1988) succinctly describe Item Response Theory (IRT) models as "the relationship between a person's level on the trait being measured by a test and the person's response to a test item or question."(p. 2745) IRT is an example of a *Latent Trait Model*.

Hambleton (1989), supported investigating the use of Item Response Theory over Classical Test Theory due to its ability to deliver on the following properties: Statistics generated are not group dependent; Results describing the ability of students are not dependent on the difficulty of items comprising the test; Models can be generated that enable matching of test items and student ability; The model can go beyond the true-score assumption; Parallel tests to establish reliability are not mandatory; Ability on a single trait can be measured in relation to the items comprising the test; The statistical procedures enable "estimates of different standard errors of measurement for students of different ability levels" and "goodness-of-fit investigations" (p.149), can be carried out in relation to students and test items.

Izard (2000) and Griffin (1998) support the use of Item Response Modelling as it provides data that can be interpreted in a diagnostic manner plus monitoring the progress of groups and individuals. Item Response Modelling also enables a variety of combinations and comparisons to be made in a valid manner. *Classical Test Theory* (CTT) data relating to test items and students could not be compared easily.

Interpretation and analysis should occur throughout the learning process. Item Response Modelling procedures are able to emphasise the importance of following student learning in a manner that enables interpretation and diagnosis of what students know and what students need to be taught. Results are reported in terms of the content and skills that are needed to be learned, not in terms of rank order within class or similar comparisons.

2. Why Choose Published Tests?

Published tests offer teachers the ability to "look before they try". The test can be evaluated against the requirements of the teachers in a school. Published tests enable teachers to be involved in the areas where they are experts (teaching and learning) and this complements the abilities of the test development experts. Published tests offer teachers the opportunity to become involved in the collection of assessment data that are relevant to the teaching and learning requirements of their class. The purpose for which the teacher wishes to collect data will relate directly to the selection of an appropriate test.

The choice of published test in an assessment program needs to satisfy the four elements identified by Lai and Griffin (2001).

Observation design - The quality of the items comprising the test used in this research.

Measurement design - The ability of the process to assign a score to individuals within the study.

Assessment - Engaging in procedures that aim to gather and interpretation of evidence regarding achievement, and

Evaluation- Judgement based on the information and the interpretation for learning.

The move from the practice where assessment on tests resulted in a one-dimensional ranking linked to the number of correct responses and a move to support for the learning needs of individual students/ groups/cohorts is now available to teachers through the use of Item Response Modelling procedures.

Turnbull (1981) maintained that standardized achievement tests, as measurement tools deserve recognition as a powerful tool in education but suffered in esteem due to the set of

qualities others had claimed they could attain. The "need to assess how much students have learned has been fundamental in education for as long as there have been students and teachers." Prior to the development of standardized achievement test the exacting business of assessment was based only on teacher judgements. Teacher judgements were based on observation, conversation, work samples, quizzes and teacher-developed tests. Turnbull reiterates, "All these bits of information entered into the marks the teacher gave. They still do and they should." He also states that standardized achievement tests "held promise of improving their information" available in the exacting business of marking. The natures of these improvements in information are; consistently *accuracy* results, removal of teacher's *objectivity* issues and *comparability* within groups of students. Turnbull summarizes: " We had, then, the basis for a fine combination of techniques: standardised achievement tests, which could measure sheer accomplishment in several areas very well, and teacher judgement, which could add dimensions inaccessible to standardized achievement testing but pertinent to the interpretation of pupil scores." (Turnbull, 1981, p.1)

This new development, *Standardised Achievement Testing*, was adopted enthusiastically but this resulted in misinterpretation and poor use of test scores. Turnbull identified three fallacies: "... testing has suffered ... from the excessive expectations of its most devoted advocates ...". In his opinion, *Standardised Achievement Testing* devotees " ... invested test scores with a precision they never possessed, read into achievement test scores much more than they really tell" and "expected the test to compensate ... for the differences in academic development of children whose learning opportunities have differed dramatically."

The most serious weakness was that the results were interpreted by comparing students with each other (losing sight of their curriculum achievements) and describing learning in terms of status relative to one or more other groups. Information about what individual students could do or not do was largely ignored.

1. How often to test with published tests?

The frequency of assessment is linked to its purpose. The purpose of an assessment program using published test at the start of the school year is to provide current information to a teacher about their new grade of students. This information can be used to support the development and delivery of curriculum matching the teaching and learning needs of the grade. "If test results are to be used to inform teaching, then they need to be administered early in the year." (Forster, 2001)

Frequent assessment of learning may not necessarily measure an increase in learning. If the teaching-learning conditions are not addressed then the repeated assessments will merely confirm that learning has not occurred.

2. What curriculum areas to target with the testing program?

In Victorian State Government Schools there are eight key curriculum areas to be covered in each year of schooling. (English, Mathematics, Science, Technology, Studies of Society and the Environment, Health and physical Education, The Arts and Languages other than English) The assessment strategies already used in the school would be strengthened and enhanced by the use formative tests in the core curriculum areas English and Mathematics. Curriculum areas that employ creativity would not be suitable.

3. Problems teachers face in a school using assessment models.

In 1998b, Black and Wiliam (p.11) concluded a review of current literature regarding *Assessment by Teachers*. They identified three key problems relating to teachers using assessment.

- *That formative assessment is not well understood by teachers and is weak in practice;*
- *That the context of national or local requirements for certification and accountability will exert a powerful influence on its practice; and*
- *That its implementation calls for rather deep changes in both teachers' perceptions of their own role in relation to their students and in their classroom practice."*

These features have implications for research in this area. Research, which simply interrogates existing practice, can probably do little more than confirm the rather discouraging findings reported above. To be productive therefore, research has to be linked with a programme of intervention. If such intervention is to seek implementation with and through teachers in their normal classroom, it will be changing their roles and ways of teaching; then the formative initiatives will be part of a larger pattern of changes and its evaluation must be seen in that larger context. More closely focused pieces of research might have to be use imported researchers because teachers cannot be expected quickly to abandon habitual roles and methods for a limited experiment. Thus at least some of the research that is needed will inevitable lack ecological validity.

Black and Wiliam (1998b, p.1) describe formative assessment as "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged." Assessment can be a time consuming activity for teachers. In schools we need to become more effective at balancing the assessment of students in a manner that ensures quality teaching and learning can occur. Assessment models need to offer reliable, relevant and quality feedback to teachers, students and parents.

4. Steps taken to overcome problems.

The need to develop achievement data that are useful to teachers in improving the learning opportunities for our students has been our incentive. An essential element of the program from conception has been the location and involvement of individuals that have had the expertise to guide the school through the changes that were required to enable successful shift in attitudes. Along the way the teaching and leadership staff at the school have gained new knowledge and skills.

The initiative arose from the Principal's concern regarding the monitoring of student achievement in a manner that produced improved teaching opportunity for teachers and was supported by accurate identification of the learning needs of students.

From the onset it was recognised by the school's leadership group that the active and supportive involvement of all teaching staff was desired. Leadership developed a strategic plan and this evolved into a general plan that was based on teacher-owned assessment and teaching based on the information obtained from the assessment tools

4.1 Long term planning, responsive to student, teacher and school need

The timeline for these developments shows the detail required to effect change.

1995 Funding and support of School Council

- Establish a budget including professional development, test purchase, marking and data interpretation
- Initial timelines developed
- Outside experts and critical friend sourced

1996 Initials Program developed

- Involvement of all staff from onset
- Professional development built into all stages

1 day off site (What is in a Published test package? What can they tell us about what we need to teach?)

The development of knowledge and understanding by all in relation to tests.

Ordering of many published test teacher wished to check out

Teacher discussion and selection of a battery of tests they wished to trial via the use of C. Crofts' Test Evaluation Sheet (1980)

Test levels to be trailed at different grade levels

- Battery of test include teacher designed tests
- Testing procedures - Class teacher to administer as then able to be responsive to student anxiety if it occurs.
- Mid-year Battery of tests purchased and administered to students
- Tests marked externally and information returned to teachers
- Analysis of spreadsheet data by teacher, an initial learning phase, grouping and discussion

1997 Program Refinement

- Battery of tests modified. Teachers discuss criteria for test selection

A and B formats tests preferable, Range covers from age 6 to 11 years

Teachers selected the best test formats available

- Core of tests selected and locked in for the next three years
- Baseline data developed
- Equating process carried out in relation to tests
- Teachers start to move away from total score information
- Australianization of terms and pictures in selected measurement tools

Diagnostic Spelling Test by D.
Vincent and J. Claydon

Mathematics 7-11

PRS Pty Ltd, Melbourne, Australia, 1997

Effective Reading Test

D. Vincent and M. de la Mare

NFER-Nelson Publishing Company, UK, 1995

- Professional Development

Moderation against the Victorian
Curriculum Standards
Framework documents started

Grouping Strategies in the
classroom

Approaches to learning and
teaching strategies

- Testing moves into Term 1 of the school year to maximise teaching and learning opportunities
- Information sharing with outside providers and critical friend

1998 Program consolidates

- Testing occurs in second and third week of school year
- Double marking and data returned to school prior to Easter break
- Continued financial support of School Council
- Review of curriculum strengths and weakness carried out by teachers and leadership
- Professional development continues

Learning styles of students

Procedure review

Areas identified as needing
development for teaching staff -
calculators

Development of inside expertise

Moderation against the Victorian
Curriculum Standards
Framework documents continued

- Learning Improvement plans developed for 30% of students in each class. Parameters of 20% from lower achievers and 10% of high achievers
- Profiles developed by outside expert and used to follow student through year levels in mathematics and Spelling

- Teachers provided with one hour of time, *Special Needs* to work with identified groups and/or individuals in grade
- Timetable developed to support *Special Needs* program
- Information sharing with outside providers and critical friend

1999 to 2002 Program continues to develop and respond to teacher and school needs.

- Ongoing financial support of School Council
- Professional development of one day for result analysis with the support of developing in house expertise and outside provider
- Professional development for teachers in data analysis and *Item Response Modelling (IRM)* variable maps
- Moderation against the Victorian Curriculum Standards Framework documents completed
- Increase to two hours of *Special Needs* time allocated to classroom teachers to support effective teaching and learning
- Answers found to questions and queries by teachers
- Use of mathematics profile for the **Mathematics 7-11** test
- Use of spelling profile for the **DST Spelling** test
- Monitoring of teacher contact across the whole school and curriculum areas of English and Mathematics
- Use of "B" format and start of the year test to provide teacher feedback on effective teaching and learning
- Review of selected tests
- Checking that valid information is still being provided by the tests using IRM procedures

4.2 Timetable

The school's timetable became an essential element of the program. The school offers Specialist subject areas of Art, Music, Physical Education and Library. Students who stay to work with their teacher must attend the Specialist area on the same day as their class.

For example: Teachers are enabled to keep a small group of their students (with an identified learning need) with them for an hour of teaching and learning.

	Grade 3C	Grade 3L
Art Specialist Area	Grade goes to Art + 5 students from 3L - 5 students	Grade goes to Art + 5 students from 3C - 5 students
Teacher's Purpose	5 students work on addition	5 students work on place-value

	with own teacher	with own teacher
--	------------------	------------------

The teachers have developed a common understanding, no student is to miss out on specialist curriculum areas. Grade and Specialist curriculum area teachers' work together ensuring that learning opportunities are maximised for all students. In this way all teachers have become involved.

The *Special Needs* teacher's timetable operates before lunch, when students and teachers are fresher. The *Special Needs* teacher works in the classroom while the teacher works with individuals close by.

4.3 Professional development

Professional development has been on going and touched all teaching and leadership staff. When possible professional development is on location or close to the work place. This development had a direct relationship to needs expressed by teachers or arisen for analysis of results in relation to curriculum provision. Interesting issues have arisen but are addressed in an open manner and answers sought.

For example several teachers felt that the proof reading passage was presented in cursive script. The Grade 3 students have just started to move into cursive writing. Did this factor affect the results they achieved? Action: As 'A' and 'B' formats of the Spelling test were available we were able to type up the second passage in infant script. The second proof reading passage was given to the students two weeks after the other. It was also marked and the data compared. No change in results showed. Both passages behaved in the same manner in relation to the students.

4.4 Timing of testing

Testing is scheduled to occur at the start of the year, as its intent is to help teacher identify student curriculum and learning requirements quickly. These students will be with them for the year ahead. This has enabled the program to keep a positive and forward focus.

4.5 Learning Improvement Plans

The school has a prescribed a set of criteria that needs to appear on a students Learning Improvement Plan but teachers are free to design their own or use the school pro-forma.

The criteria are as follows.

- specific achievable goals in a curriculum or social area
- the plan of action the teacher is going to take
- parent in attendance at a meeting or phone contact with a copy of plan signed and returned to the class teacher
- suggestions for supportive action at home

(For more detailed information see Izard et al, 1999, Izard 1998, Silis & Izard 2001.)

4.6 Resources

Resources are responsive to needs identified by the program early in the year enabling teachers to deliver the required teaching and learning. Three forms of resources are

identified human, physical and disposable. The human resource equivalent to two days of a teacher is funded to enable teachers to receive the two hours of *special needs* teaching time with their students. The remaining hours are provided via the timetable and involve the Specialist subject teachers at the school. Physical resources such as equipment calculators, computer programs, educational aids and readers are provided and the disposable materials that are consumed in the teaching and learning process.

4.7 Curriculum and student monitoring

The teachers have selected the tests in use and have modified the range of tests used over the time the program has evolved. The tests have been able to identify areas of learning needs in all students. The Item Response Modelling procedures provide teachers with data that easily converts to useful information for teaching purposes. A picture of the students within grade groups is able to maximise and support small group and individual learning opportunities for students and teachers.

The measurement tools have been mapped onto the Curriculum Standards Framework, the Department of Education's curriculum guide. This supports teachers in meeting accountability recommendations. The individual is catered for as the test information identifies learning that students require. This enables teacher to take action. A student's position can be viewed in two ways. The variable map locates them in relation to the test item difficulty and their peers. The assessment occurs at regular intervals and progressive improvement can be traced. In mathematics and spelling a profile has been developed. The profiles track learning over the years of schooling. Learning as a process over time comprises a collection of skills that builds ability. The *effect-size* of the changes in cohort groups can be calculated. This enables reflection and required actions to be planned.

Shinn and Good, in 1993 established six features central to curriculum-based assessment. For an assessment model to be effective these features need to be accounted for by a school. On reflection it is pleasing to see that they have been addressed by the Testing for Teaching program.

- "Assessment exercises should faithfully reflect the main learning aims and should be designed to evoke evidence about learning needs.
- The main purpose for assessment is formative.
- Validity is paramount-seen as ensuring that instruction decisions taken on the basis of assessment evidence are justified
- The focus of attention is the individual learner and individually attuned remedial action.
- The information from assessment should serve to locate the individual's attainment in relation to criteria for learning, but that this location should also be informed by norm data on the progress of others working to the same curriculum.
- The assessment should be frequent so that the trajectory of learning over time can be traced: the gradient of learning success is the key indicator-to follow each pupil's progress in general, and to indicate cases of special need."

Time has been allowed to enable changes in practice to occur. "Where changes have been introduced with substantial training or as an intrinsic part of a project in which teachers have been closely involved, the pace of change is slow because it is very difficult for teachers to change practices which are closely embedded within their whole pattern of pedagogy." (Black and Wiliam, 1998a, p.10)

Review procedures are carried out in Term 1 of the school year when concerns are fresh in the minds of all concerned. Plenty of time is available for any changes that need to be in place for the following year.

5. The Information provided to teacher

After students have completed tests they are collected and sent away for double marking by an external provider. The provider used by the school is PRS. Computer disc and print out are provided to the school.

Grade teachers receive the information about their class in alphabetical and rank order lists. Sub-scale scores are show in the following areas.

Mathematics 7-11

- Non Numerical processes
- Understanding Number
- Computations And knowledge
- Mathematical Interpretation
- Mathematical Applications
- Levels M10 and M11 information on calculator usage

DST Spelling

- Homophones
- Common Words
- Proof-reading
- Letter Strings
- Nonsense Words
- Dictionary Use

ERT Reading

- Using relationships within text
- Acting upon text
- Employing reading strategies appropriate to text and purpose
- Making an affective imaginative or personal response to reading
- Critical awareness and evaluation
- Location and selection skills

The results are placed into the format of Figure 1 below. Figure 1 is an example of the information provided to a class teacher. Information is also provided in variable map format (See Figure 2).

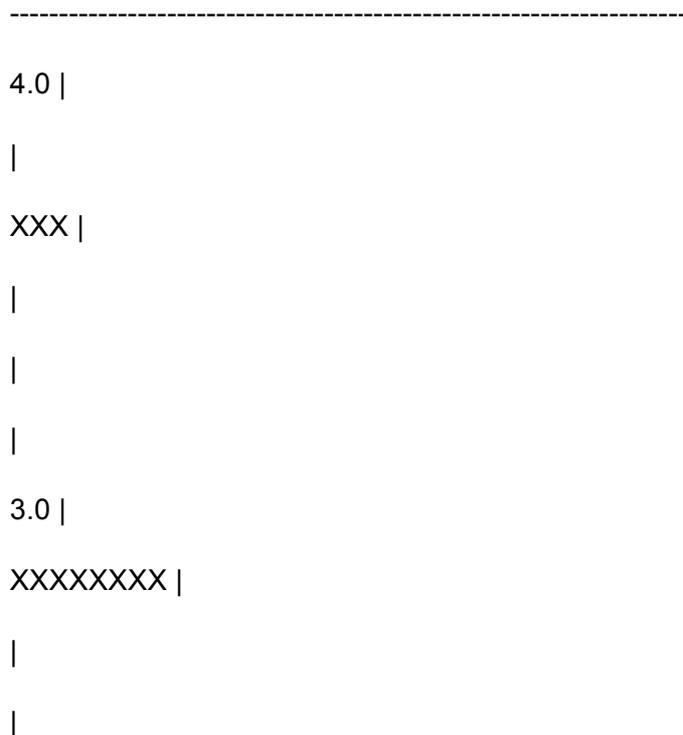
Grade H3						
Student ID Code	Non Num Process	U/stand Number	Comput & Know-ledge	Maths Interp	Maths Applic	Total

	N=5	N=9	N=10	N=8	N=8	N=35
SSS	1	4	2	1	4	12
KLJ	4	5	6	3	3	21
MKI	2	7	7	4	7	27
-----	-----	-----	-----	-----	-----	----
NHY	5	5	8	5	7	30
Mean	2.69	4.08	5.23	4.77	1.54	18.3
Std Devn	0.99	1.59	2.39	1.58	1.08	5.82

Figure 1: Example of a Class List for a Mathematics 7-11 Test

Item Estimates (Thresholds)

(N = 29 L = 19 Probability Level=0.50)





| 3

XX |

2.0 |

|

XXX |

| 9

|

XX |

| 5

X |

1.0 | 18 19

XXX | 4 10

|

| 8 14 15

XXX |

|

XX |

0.0 |

| 13

|

| 2 17

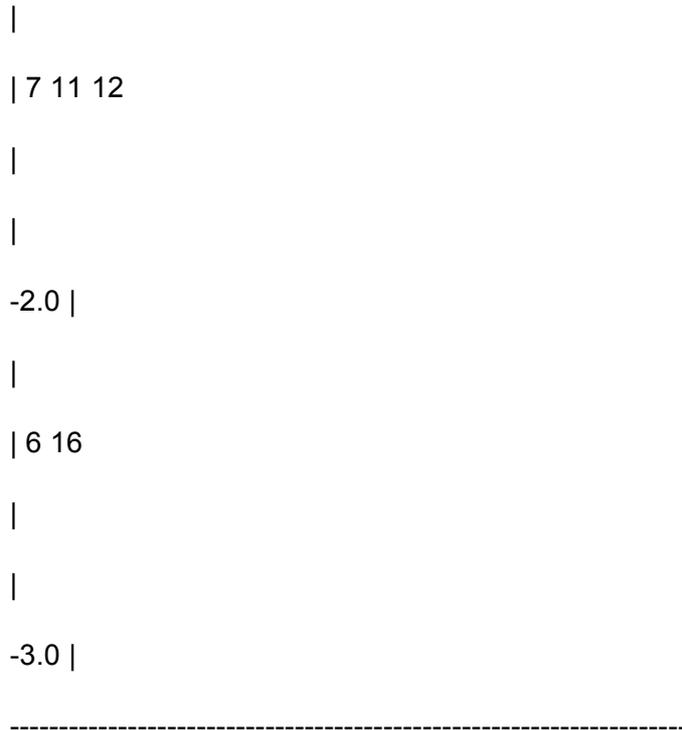
|

| 1

X |

|

-1.0 X |



Each X represents 1 student

Figure 2: Example of Variable Map

5.1 How the teachers use the information.

The information provided by the testing is used by the teachers to grasp a view of their grade in the curriculum areas of English - Reading skills, Spelling and Mathematics. As the students in the class are new to the teacher this information was not available previously.

Lai and Griffin, 2001, state that; "The manner in which tests helped teaching and learning depends on whether appropriate information could be derived from the tests and whether correct interpretations could be made and communicated to others based on this information." (p. 2)

The variable map (Figure 2) enables teachers to see how students have responded to each item comprising the test. The variable map is connected to a teacher's classroom of students and each class forms its own patterns. The items the grade found difficult appear at the top and the easier at the bottom. 0.0 is the point of average difficulty for the grade group. The teachers can relate each question back to the relevant areas of the curriculum. This provides teachers with a picture of content areas where the class needs further consolidation or teaching.

Test-item data expressed in sub-scores provides to the teacher/curriculum planner an improved position for planning the learning needs of students. All teachers are curriculum planners. Teachers need support in identification of the needs for each grade they teach. Curriculum frameworks are compiled of *outcome statement* a guide to what is expected to be covered by teachers. The sub-score and *variable map* provided by the program support teachers in carrying this out.

As the program has developed the outside experts have been able to support teachers in this process of improving the learning opportunities for students by developing profiles in mathematics and spelling. The profile forms show progressive achievement in the sub-score areas over the schooling of each child. Valid comparison can be achieved in relation to this progress because the profile charts make due allowance for errors of measurement.

Areas of strength and weakness can be identified across the whole class as groups or on an individual basis. Individual Learning Plans are developed and discussed with the parents of students. Learning Improvement Plans outline areas that need to be taught in the classroom and will be covered in the classroom and *Special Needs* time. Activities and strategies are provided to parents for work on at home.

During the initial years of the program only teachers looked forward by analysing and planning for the current year's group of students. Now many ask to see the results achieved by last year's class group, and in particular, the students who have been the focus of *Special Needs* teaching and learning time.

5.2 What we found out by using assessment data.

- Teachers found out what learning had been held by students over the long summer break.
- Early identification supported more time for teaching. When the data is available very early in the school year teachers do not feel compelled to use teacher-designed tools that contain internal validity problems. (See Silis & Izard, 2001)
- Teachers have learnt how to use assessment data generated from test to improve knowledge about the student in grades.
- Areas of the curriculum not being covered very well were identified supporting professional development and resources investment by the school.
- The transfer of assessment data onto profiles enables a strong visual link to the achievement pattern of a students learning to be celebrated by teachers and parents.
- Curriculum planning developed a stronger link to the actual needs of the students.
- The identification of strengths and weakness supports sound curriculum planning.
- Item Response modelling (founded on Rasch measurement) procedures are able to provide teachers with worthwhile information about students
- Department accountability measures can be answered.
- That as a whole school staff an idea can evolve into an initiative that involves all and results in changes to the school in relation to teaching and learning.

The program has addressed the key weakness identified by Black and Wiliam

in their 1998 (p.9) review of classroom assessment practice.

- *"Classroom evaluation practices generally encourage superficial and rote learning, concentrating on recall of isolated details, usually items of knowledge which pupils soon forget.*
- *Teachers do not generally review the assessment questions that they use and do not discuss them critically with peers, so there is little reflection on what is being assessed.*
- *The grading function is over-emphasised and the learning function under emphasised.*
- *There is a tendency to use a normative rather than a criterion approach, which emphasises competition between pupils rather than personal improvement of each. The evidence is that with such practices the effect of feedback is to teach to weaker*

pupils that they lack ability, so that they are de-motivated and lose confidence in their own capacities to learn.

- *Teachers appear to be unaware of the assessment work of colleagues and do not trust or use their assessment results."*

6. Conclusion

For assessment practice to change in schools teachers must be given the opportunity to be involved in the change. Adequate time must be planned into the process. The pivotal role that teachers play must be acknowledged if leadership (top down) or superficial activity is to be avoided.

The assessment system has helped teachers to deliver learning that is closer to the specific needs of students. Teachers are still using a range of assessment strategies to complement the information provided at the start of the school year by the testing. Teachers are in a better position to base learning on what students already know without interference from personal bias and other emotional issues.

With the development of the profile forms progress from year to year is shown for each student in a form that teachers and parents can understand. This has enabled the celebration of success and targeting of students for further investigation into specific learning problems.

Curriculum is adapted to meet the individual student needs of grades of students in Mathematics and English-Reading. A powerful and focused learning environment has resulted.

External indicators (Triennial Review and State Achievement data) confirm internal evidence that learning has been successful. The use of outside expert help has provided a sound statistical base to assessment measurement. The 'effect size' or 'added value' to cohorts of students in relation to learning in the core curriculum areas of Mathematics and English can be demonstrated.

The program will continue to have a whole school approach involving all stakeholders in the process (teachers, students, leadership and parent). Review and modifications will continue to occur as they form an essential ingredient resulting in its success. Time will be allowed for change to occur.

References

Adams, R.J. & Khoo, S.T. (1993). *The interactive test analysis system*. Hawthorn, Victoria: Australian Council for Educational Research.

Black, P. & Wiliam, D. (1998a). "Assessment and Classroom Learning," *Assessment in Education*, Vol, 5, p.7-74.

Black, P. & Wiliam, D. (1998b). " Inside the Black Box: Raising Standards Through Classroom Assessment," *Phi Delta Kappan*, p. 139. (Also at www.pdkintl.org/kappan/kbla9819.htm) 22:11:02, 3.56 p.m.

Board of Studies. (2000). *Achievement Improvement Monitor: English and Mathematics: Testing Component Year 3 and Year 5, 2000 Reporting guide*. Carlton, Victoria: Board of Studies.

Board of Studies. (1996). *LAP Data Reporting Guide 1999*. Carlton, Victoria: Board of Studies.

Board of Studies, Data Analysis Service. (1998). *Reporting Guide, Learning Assessment Project*. Carlton, Victoria: Board of Studies.

Croft, C., (1980). The Foundations of School Testing. In SET. Research Information for teachers, Item 9, No. 2, 1980. Australian Council for Educational Research. (ACER.)

Department of Education, Employment and Training, (2001). *Standards and Accountability: Quality Assurance in Victorian Schools, Benchmarks, 2000*. Melbourne, Victoria: Department of Education, Employment and Training.

Department of Education and Youth Affairs (DETYA), (2000). *Numeracy, A Priority for All: Challenges for Australian Schools: Commonwealth Numeracy Policies for Australian Schools*. Canberra: Commonwealth of Australia.

Forster, M. & Masters, G. (1999). *Paper and Pen: Assessment Resource Kit*. Melbourne, Victoria. Australian Council for Educational Research

Forster, M, (2001) A Policy Maker's Guide to, *System-wide Assessment Programs*, ACER

Lai, K., & Griffin, P., (2001, December 2-6). *Linking Cognitive Psychology and Item Response Models*. Paper presented at the 2001 annual conference of the Australian Association for Research in Education, Perth, Australia.

Lord, F. M., & Stocking, M. L., (1988) Item Response Theory. *International Encyclopaedia of Education Research and Studies*, (p. 2745-2748). Oxford, Pergamon Press.

Hastings, J., T., (1974). *Curriculum Evaluation: Commentaries on Purpose, Process, Product*. Curriculum Evaluation: The Why of the Outcomes. USA, D.C. Heath and Company

Hambleton, R. K., (1989). Principles and Selected Applications of Item Response Theory. In R.L. Linn, *Educational Measurement: Issues and Practices*, (3rd Ed.). (p.147-200), New York, American Council on Education.

Griffin, P., (1998), Australian Association for research in Education Conference: *AA Twenty year Equating Study of mathematics Achievement*. Adelaide. AARE, 1998.

Izard, J. F. (1991). *Assessment of learning in the Classroom*. Deakin University: Deakin University Press.

Izard, J. F. (1992). Patterns of development with probability concepts. In M. Stephens & J. F. Izard, (Eds). *Reshaping assessment practices: Assessment in the mathematical sciences under challenge*. Hawthorn, Vic.: Australian Council for Educational Research.

Izard, J. F. (1994). Strategies for assessing projects and investigations: Experience in Australia and United Kingdom. In Mauritius Examinations Syndicate (Eds.). *School-based and external assessments*. (p. 214-222). Mauritius: Mauritius Examinations Syndicate.

Izard, J.F. (1998a). Validating teacher-friendly (and student-friendly) assessment. In D. Greaves & P. Jeffery (Eds.). *Strategies for intervention with special needs students*. (p. 101-115). Melbourne, Vic.: Australian Resource Educators Association Inc.

Izard, J. (1998b). *Assessment of Cognitive Skills*. National Assessment Institute. Melbourne: University of Melbourne.

Izard, J., Jeffery, P., Silis, G.F., & Yates, R.L., (1999). Learning Disabilities: Advocacy and Action. In Peter Westwood & Wendy Scott: (Eds.) *Testing for Teaching Purposes: Application of Item Response Modelling (IRM) teaching-focussed assessment practices and the elimination of learning failure in schools*. (p. 163-188). Australia. Australian Resource Educators' Association Inc. (AREA)

Izard, J., (2001). Consultancies and Research Projects in Asian Countries. Paper presented at the Research Conference for Postgraduate Students & RMIT Staff. 1st June, 2001, Melbourne, Australia.

Ministerial Council for Education, Employment, Training and Youth Affairs. (MCEETYA), (1999). *National Report on Schooling in Australia: Special focus in this issue on: Literacy, Numeracy, Education of Indigenous Students*. Curriculum Corporation, Melbourne, Australia.

Piper, K., (1997). *Riders in the Chariot: Curriculum Reform and the National Interest 1965-1995*. Melbourne: Australian Council for Educational Research. (ACER)

Professional Resources Services, (1997-2001) Mathematics 7-11. & DST Spelling. Melbourne. P.O Box 71, Coldstream, Victoria, 3770.

Rowe, K. J., (1999). Assessment, Performance Indicators, 'League Tables', 'Value-added' Measures and School Effectiveness?: Consider the issues and 'Let's Get Real! Centre for Applied Educational Research, The University of Melbourne.
[<http://www.area.edu.au/99pap/row99656.htm>] 22:11:02, 3.35 pm.

Shinn, M.R., & Good 111, R. H., (1993) CBA: an assessment of its current status and prognosis for its future, in: J.J. Kramer (Ed) *Curriculum-Based Measurement*, p 139-178, Lincoln, NE, Buros Institute of Mental Measurement.

Silis, G., & Izard, J., (2001) Monitoring student progress to use a published tests or teacher developed tests. [<http://www.area.edu.au/01pap/sil01289>. 22:11:02

Turnbull, W., (1981). Achievement Test Scores in Perspective. In SET. Research Information for teachers, Item 12, No.1, 1981. Australian Council for Educational Research. Camberwell, Victoria. (ACER.)

Traub, R.E., (1983). "A Priori Considerations in choosing an item Response Theory, (p. 57-70). Vancouver, BC: Educational Research Institute of British Columbia.

Wright, B.D, & Stone, M. H., (1979). *Best Test Design: Rasch Measurement*. Chicago: Mesa Press.

Victorian Curriculum and Assessment Authority. (VCAA), (2000). *Achievement Improvement Monitor: AIM Testing 2001 Reporting Guide, English and Mathematics Testing Component, Year 3 and Year 5*. VCAA. East Melbourne, Victoria, 3002.

Virgona, J., & Greaves, D., (2001). Insightful Assessment Through Item Response Modelling. *Australian Journal of learning Disabilities*. Vol. 6, No. 2, June.

Victorian Curriculum and Assessment Authority (VCAA), (2000), *Achievement Improvement Monitor: AIM Testing 2001 Reporting Guide, English and mathematics Testing Component, Years 3 and Years 5*, VCAA, East Melbourne, Victoria 3002.

Wiliam, D., (1996). *National curriculum assessment and programmes of study: validity and impact*. *British Educational Research Journal*, 22 (1), p.129-141.