# Chapter 10
# Teachers' Use of Assessment Data

**Patrick Griffin**

> *. . . it was important . . . to say . . . that a test signals where to start intervention and not the end point of instruction . . . the idea of Evidence Based Teaching is very important and I complement you on highlighting the idea and your emphasis of obtaining appropriate resources to implement effective use.*
>
> Robert Glaser (personal communication, 28 June 2007)

## Developmental Learning Framework

The emphasis has to be on development, and teachers need to be clear about the difference between deficit and developmental learning approaches. Clinical and deficit approaches sometimes focus on the things that people cannot do and hence develop a 'fix-it' approach. 'It is a myth that intervention is only needed for the struggling student' (Tayler, 2007, p. 4). Developmental models not only build on and scaffold existing knowledge bases of every student, but they also have to be clinical in that they focus on readiness to learn and follow a generic thesis of developing the student. They ought not entertain a deficit thesis of focusing on and emphasising 'cures' for learning deficits. In order to become a specialist in developmental learning, teachers need to have expertise in developmental assessment because it is integral to the formulation of personalised learning plans. How often have we heard the teacher say 'we start where the student is at'? It is impressive rhetoric but unless teachers are capable of monitoring learning and identifying where both the student and the teacher are 'at' on developmental pathways, and targeting intervention, it is likely that the rhetoric may be realised only serendipitously. This chapter examines an overall approach to the use of assessment data to inform teaching intervention decisions and then illustrates the possible results that can be achieved.

In a developmental framework there is a need to break the link between whole-class teaching and instructional intervention. Teachers have to focus on 'individual developmental and personalised learning' for every student. When teachers pursue a developmental model, their theory of action and psychology of instruction needs to

P. Griffin (✉)
Melbourne Graduate School of Education, The University of Melbourne,
Parkville VIC 3010, Australia
e-mail: p.griffin@unimelb.edu.au

C. Wyatt-Smith, J.J. Cumming (eds.), *Educational Assessment in the 21st Century*,
DOI 10.1007/978-1-4020-9964-9_10, © Springer Science+Business Media B.V. 2009

focus on theorists who have promoted and given substance to developmental learning. Being able to identify the ‘Vygotsky zone of proximal development (ZPD)’ is fundamental to the identification of where a teacher would intervene to improve individual student development (or ‘where the student is at’). Teachers need to be able to recognise and use the evidence to implement and monitor within the Vygotsky approach. Which developmental theory underpins the work is negotiable, but choosing a developmental theoretical basis is an important aspect of all forms of teacher education (both pre-service and in-service) if teaching for individual developmental learning is to be realised.

It is also evident that when a developmental model of learning is implemented, the teacher has to reorganise the classroom and manipulate the learning environment to meet the needs of students. Manipulation of the learning environment is an important skill, and the way in which a teacher links classroom management, intervention strategies and resources used to facilitate learning is always a challenge. The strategies need to be guided by a developmental framework of student learning.

## Changing the Paradigm: Assessing ‘What’ and ‘How Well’ in Learning

The topic ‘assessment’ still conjures images of tests. Tests conjure ideas of standardised measures of literacy and numeracy and ‘easy-to-measure’ disciplines. Standardised measures conjure normative interpretations, labelling, ranking and deviations; there is a widespread belief that ease of measurement dictates assessment and that the hard-to-measure subjects are ignored. Assessment and measurement are in turn seen as reducing learning and curriculum to what is easy to measure. In fact, nothing is too hard to measure. As Thurstone (1959) said, ‘If it exists it can be measured and if it can’t be measured is doesn’t exist’. It all depends on how measurement is defined.

It is not necessarily true that only easy areas are measured. A slight reconceptualisation of measurement can allow assessment to focus on difficult areas to measure and help link learning to targeted intervention. Educational measurement typically demands technical skills, and its specialists are generally engaged in large-scale testing programs at systemic, national and international levels. Assessment, on the other hand, requires a different but overlapping set of skills and is linked more generally to teaching and intervention, although measurement can and should be conceptually at least underpinning the assessment. Too often at the school level, or in teacher education, measurement or technical aspects of assessment are seen as encroaching on discipline areas of curriculum. It is often regarded as a subdomain of curriculum. Of course, assessment is a part of curriculum, but it needs explicit treatment and the development of the relevant skills base.

Griffin & Nix (1990) defined assessment as the process of observing, interpreting and making decisions about learning and intervention, whereas measurement was regarded as the process of assigning numbers to observations. Neither of these is curriculum. *It is only when the numbers have a meaningful interpretation that measurement and assessment begin to merge and they build a link to curriculum.* The bodies of knowledge for measurement and assessment are different but overlapping. What a psychometrician does is not what a classroom teacher does, but the logic and framework that a psychometrician works with can be used to inform classroom practice and, where it is, the teacher is offered a more rigorous approach to personalised and clinical approaches to intervention.

In a curriculum framework, teachers are taught to identify what is wrong, mostly using test items or assessment tasks (rich or otherwise) to identify what the students cannot do and then to concentrate on fixing, or curing, the problem. The focus on fixing deficits is the 'norm'. The motive looks like 'fix the weak and let the strong learn on their own' (Stanovich, 1986). This leads to the situation in the classroom where one group of students struggles to learn things far beyond its learning readiness, another group is coasting ahead of the 'pack' and the rest of the class is being taught as a homogeneous group.

It is possible to turn it around and engage every student at their point of readiness to learn. A shift towards developmental-learning outcomes demands both a change in thinking about curriculum and developmental learning, and the method for implementing change across multiple levels of student learning.

The first step, assessment, monitors what a student needs to learn. It is not always possible, and certainly not necessary, to assess everything that all students need to learn, but assessing a good sample of the attitudes, skills and knowledge is important. Hence, there is no need to list all the discrete skills as a definitive litany of mandatory achievements that must all be demonstrated. The attitudes, skills and knowledge that students acquire are not isolated, discrete entities. They are best learned when they are conceptualised and introduced as sets of cohesive and interrelated skills, attitudes and knowledge that build to a developmental continuum.

## Test Construction and Developmental Progressions

Good tests and good assessments have a psychometric basis, in that they attempt to measure a specific developmental pathway that psychometricians call a 'variable' (or construct). Sometimes a test or assessment task might attempt to measure a small set of variables; teachers need to know and understand the nature of the underlying variable. Addressing or teaching to the underpinning construct is important because it takes away any focus on each of the individual test items.

Embretson and Reise (2000) showed that the overall test developmental variable is made up of three parts. These were the underlying developmental progression, or 'latent' variable, the items that point to the developmental progression, or 'manifest', variables and the error associated with each of the test items. Each individual item can measure a range of things, some of which are related to the latent developmental progression, but they each measure 'other things' as well; the extent to which these 'other things' influence the measure is related to the reliability of the test. The extent to which the items relate *as a set* to the developmental progression emphasises the validity of the test and of the interpretation of the construct. A test instrument is constructed to measure student ability on a developmental progression (or, technically, a latent construct), not on the individual items. In this sense the nature of the item selection for the test is, while not unimportant provided that the item set is sampled from the appropriate domain, designed to measure a specific trait. Figure 10.1 shows the relationship between the latent construct (developmental progression), the test items and the error terms, or ‘noise’, in the measurement of the construct.

Embretson and Reise (2000) also showed that the test items represented separate ideas that built into the developmental construct. It is the construct or the progression that mattered, not the specific choice of items. The items are replaceable, provided that the ‘bigger idea’ or the developmental progression can remain the main basis of the assessment and understanding of the student development. The items might be represented as ‘little ideas’ all contributing to the measure of the ‘bigger idea’ embedded in the progression. At the same time, each test item is measuring other things as well. The tendency to measure or reflect the influence of other things is known as ‘measurement error’. The danger in focusing on the ‘little ideas’ (or teaching to the test) may mean that the process is unable to see the ‘forest for the trees’.

It is possible that some items may have different properties in different schools or in different curricula. If this is the case it is identified using a process called ‘differential item function analysis’ (Adams & Khoo, 1995). When measurement

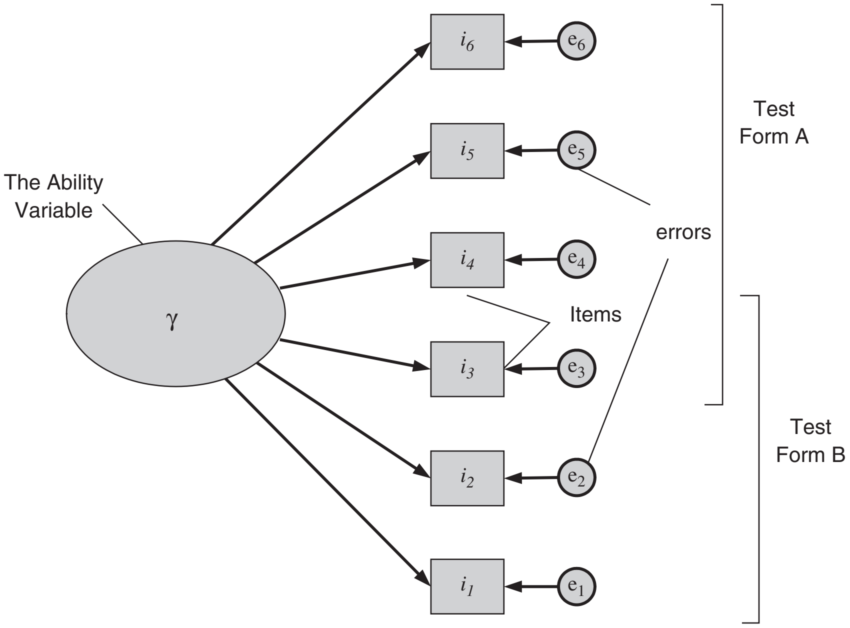


**Fig. 10.1** Relationship between items, errors and the latent variable

specialists identify this effect, the offending items are usually excluded from the test intended to measure a common construct across different student samples. In classroom tests it is possible to identify this effect with a little analysis, but the detail is beyond the scope of this chapter. It is generally the case that the standardised test used in high-stakes programs have been through this evaluation process and offending items have been removed. The sample of items left still measure the same underpinning construct but the absence of some topics often raised the ire of content-focused assessment specialists.

It is also important to remember that a test is one way to observe student behaviour in order to make an inference about their ability and their location on a developmental progression. It is just one form of evidence and generally needs to be considered along with other information that the teacher might have. Once the developmental progression is described the teacher can use it, the test data and other evidence to make decisions about the student's level of development.

Perhaps the greatest weakness of the reports linked to high-stakes accountability testing programs is not the lack of consistency across schools, but the failure to explicate the nature of the underlying variable. The reports too often have focused on individual items and encouraged this approach to interpretation. We often see media discussions of individual items illustrating student deficits and ignoring the overall picture of development or level of competence.

The competence levels are related directly to the definition of criterion-referenced interpretation of data. Glaser (1963, 1981, 1990) first defined criterion-referenced performance and development in terms of the tasks performed. However, this definition lost the idea of multiple tasks that form a cohesive and developmental continuum, and the misinterpretation of the concept in the 1970s led to the distortion of the concept. Glaser later clarified criterion referencing as 'the development of procedures whereby assessments of proficiency could be referred to stages along progressions of increasing competence' (1981, p. 935, emphasis added).

The words 'stages along progressions of increasing competence' are important in test design and calibration. However, criterion referencing is a means for interpretation rather than a means for test design, and *criterion-referenced interpretation* is the correct term rather than criterion-referenced testing. Criterion-referenced interpretation is also an excellent framework within which to use item response modelling such as the Rasch model (Rasch, 1960; Wright & Stone, 1979). Glaser's word 'stages' can cause concern because some interpret 'stages' as a strict hierarchical step. In fact, the stages are artificial divisions on a continuum. The thresholds between contiguous levels are arbitrary; a person can theoretically be placed on the continuum within each 'stage'. Progression on a continuum is *not* monotonic and individuals will make progress, but it is almost always mixed with regression depending on the context of observation.

Performance in a competency model is dependent on context. A person's performance depends on the demands made by the context and the personal factors at work at the time of the observations. These variations are generally interpreted as measurement errors, but we know that there are a range of issues that affect performance. If the test fails to engage the student, the performance will be reduced. If the student is not interested, is ill, unmotivated or in general disinterested, it will not accurately measure ability. Therefore, any use of a score to identify a level or stage is influenced by both the personal factor and issues arising from the instrument itself. Hence, the use of the term ‘stage’ does not imply a fixed description of a performance, but one that is indicative of a phase of development and different instruments or different circumstances, all of which may change this measurement. The variation in the measures is generally called ‘measurement error’. In this way, we can argue that the use of levels is more likely to capture the description of the person’s competence than a single number or score. It is educationally more meaningful (for teachers’ use) to use levels or stages of competence than a score or a number.

## *Variable Mapping*

There are ways to use the data constructively, however. Combining the ideas of criterion-referenced interpretation with item response modelling directly links a measure of the position of a person or an item on a variable (a variable map, as shown in Fig. 10.2) to an interpretation of what a student, or groups of students, can do, rather than focusing on a score or the performance relative to a percentage or a group. It also orients the use of the test data towards substantive interpretation of the measurement than does a score or grade. The combined set of procedures gives meaning to test scores and helps to establish the meaning and interpretation of the latent variable. As such, it has important implications for the validity of the variable and helps to focus attention on the ‘bigger picture’, rather than the collection of little ideas represented by the items. The little ideas represented by the items are defined by the cognitive skills that are required to get the right answer to each of the items. The process of identifying these skills is called a ‘skills audit’.

It can be seen from Fig. 10.2 that test items tend to group or cluster together at different points along an underlying dimension or scale. Once the clusters are identified, the next task is to determine whether the items within these clusters can be interpreted as having something in common. Figure 10.2 exaggerates the idea of clustering to assist in explaining the point. Each item is reviewed for the skills involved in responding correctly. The process requires an understanding or empathy with ‘how the students think’ when they are responding to the items. Experienced teachers are very good at this task and those dealing with mathematics or science instruction can readily identify the levels within the test from a skills audit of individual items.

The variable map in Fig. 10.2 shows that items are grouped according to similar levels of difficulty. Students are represented by *X*; items are represented by the circle with the item number embedded, and the interpretation of the skills involved in the item clusters is represented by the text on the right of the figure. The levels are differentiated using the horizontal lines. In this example five levels are shown, but the number of levels depends on the clusters and separation of the items on the latent variable represented by the vertical arrow. Given that the ability of the students is approximately matched to the difficulty of the items at each level, and the items and

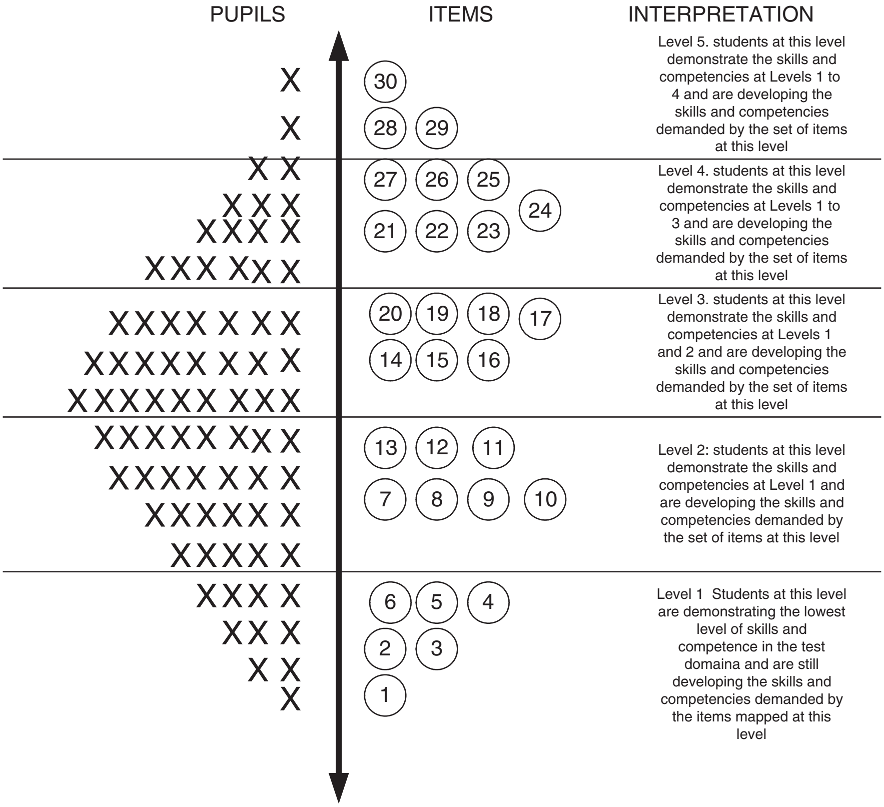

PUPILS
ITEMS
INTERPRETATION
Level 5. students at this level demonstrate the skills and competencies at Levels 1 to 4 and are developing the skills and competencies demanded by the set of items at this level
Level 4. students at this level demonstrate the skills and competencies at Levels 1 to 3 and are developing the skills and competencies demanded by the set of items at this level
Level 3. students at this level demonstrate the skills and competencies at Levels 1 and 2 and are developing the skills and competencies demanded by the set of items at this level
Level 2: students at this level demonstrate the skills and competencies at Level 1 and are developing the skills and competencies demanded by the set of items at this level
Level 1 Students at this level are demonstrating the lowest level of skills and competence in the test domaina and are still developing the skills and competencies demanded by the items mapped at this level
30
28 29
27 26 25 24
21 22 23
20 19 18 17
14 15 16
13 12 11
7 8 9 10
6 5 4
2 3
1

**Fig. 10.2** Typical variable map interpretation

students are mapped onto the same scale, the students can also be grouped within the same 'ability'/'difficulty' range as the items that have similar difficulty levels. This grouping of items (and students) identifies a kind of 'transition point' (indicated by the horizontal line), where an increase of item difficulty is associated with a change in the kind of cognitive skill required to achieve a correct answer.

When the person's ability and the item's difficulty are equal, the odds of success are 50/50. Hence, in Fig. 10.2, the students represented by the *X*s adjacent to the cluster of items have about a 50/50 chance of being able to solve the items in the adjacent cluster of items, less than 50/50 chance of solving the items above their level and a better than 50/50 chance of solving the items below their level. It is also possible to describe their level of ability by identifying the cognitive skills in the items at each level or cluster. If the student were to improve a little, they would have a better-than-even (50/50) chance of succeeding on items in the adjacent group, and it could be argued that the main task of a teacher is to increase the odds of success in each of these competency levels to a level greater than 50/50. It is also a clear identification of where the student is 'ready to learn' and will learn with assistance more readily than they would learn alone. This level where intervention is best targeted has the same interpretation as the ZPD (Griffin, 2007).

Improvement through intervention at the ZPD can take the student close to, and perhaps past, the transition point between clusters of items; the students are beginning to exhibit a change in cognitive skill. This is, in turn, defined by the set of cognitive skills demanded by the group or cluster of items. Curriculum and teaching specialist panels (or teachers) need to undertake the content analysis of the skills/competencies required to succeed on the set of items. A change in the required cognitive skill level could be directly translated into an implication for a change in teaching, and so discussions with curriculum specialists would be needed to identify the kind of instruction needed to help the student progress on the variable or construct. A summary description of these skills can then be assigned to each item and the student.

The item grouping can be justified on statistical and conceptual grounds if the items have behaved in a cohesive manner that enables an interpretation of a variable underpinning the test. This item behaviour is sometimes described as a Rasch-like manner because it is also a requirement of the Rasch model analysis. Labelling the skills is based on conceptual rather than statistical grounds. If the items within a group do not suggest a meaningful and unifying set of skills or competencies, the set may need to be ‘adjusted’ to make the interpretation clearer; that is, some items may need to be omitted because, despite statistically appropriate qualities, they may not be conceptually relevant to the underlying construct or to identifiable and comprehensible levels within the construct. This is a more powerful reason for omitting items from a test than a misfit analysis. Under these circumstances, they might not belong in the test at all. These procedures can, at times, also identify gaps in the item set. These approaches have been explained in detail by Griffin (1998). The labelling or clustering interpretation is not just a list of the skills for each of the items included in the cluster. Identifying and naming the level description involves a similar process to that used in interpreting a factor in a factor analysis. In Fig. 10.1, it is the description of the construct, or the ‘big idea’, underpinning the set of items in the cluster. This is a generalisation of the level description, treating the items in the cluster as a sample of all possible items that could represent this level of development.

There is a further advantage to this procedure. If the content analysis ‘back translates’ to match or closely approximate the original hypothesised construct used to design and construct the test, it can also be used as evidence of the construct validity. When this is linked to the index of item separation there are two pieces of evidence for the construct validity of the test (see Wright & Masters, 1983; Griffin & Nix, 1990). The technique of ‘levels’ has been used sparingly but has emerged in several international studies, including the PISA and Southern and Eastern African Consortium for Monitoring Educational Quality studies (Murimba et al., 2002). Greaney and others used the procedure in their report on the Pakistan *Education For All* project (Greaney, Khandker, & Alam, 1990), in which they cited Griffin and Forwood’s (1990) application of this strategy in adult literacy. More recently, Murimba et al. (2002) illustrated the competency levels identified in the SACMEQ tests across 15 southern African countries.

From a curriculum point of view, the clustering and labelling enable teachers to identify where targeted intervention is warranted for individuals. Systems can use aggregated data on distributions across levels for the identification of priority areas at which to thrust resources. It is a more efficient approach than that which focuses on individual items and encourages 'washback' based on item-level skills or 'little ideas'. However, recall that the skills embedded in the items represent only a sample of the population of possible skills that make up the developmental progression. The skills used in the test might only be used because they are relatively easy to obtain in large-scale testing programs. When teachers use this test information, they need to be encouraged to supplement with other observations.

## The Fundamental Role of Assessment

In any assessment, it is only possible to use directly observable behaviour because these act as pointers, or indicators, to an underpinning generalised learning. Humans can only provide evidence in the form of what they *write, make, do* and *say* (Griffin, 1997), and it is from these four observable actions that all learning is inferred. This is the basic and fundamental role of *assessment*—to help interpret observations and infer learning. The more skills are observed, the more accurately generalised learning can be inferred. Hence, there is a need to document the discrete observable skills and find a way to blend them into cohesive evidence sets so that strategic intervention can be a part of teaching and learning. The range and quality of the data (records of observed skills) enable the inference of a developmental progression, and this enables a *generalisation* to be made that can be independent of the specific set of discrete skills observed. Generalisation helps to identify where scaffolded intervention can occur (Vygotsky, 1986) for every student—high achievers as well as the lower achievers—and this is the basis of developmental learning. It eschews deficit thinking.

Of course, every student is different and no one follows a generalised pattern exactly, but it is possible to identify the typical *generalised* developmental path. It gives the teacher a framework within which to work, but it never replaces the judgment of the teacher about where to start, how to proceed or how to teach. Rather, it becomes the framework within which teaching decisions can be made (Griffin, 2007). The developmental progression is therefore an organising framework for communication, reporting and scaffolded intervention purposes. It is not a measure of performance. It is not a score, not a grade and not an assessment instrument.

## Formative Assessment and the Role of Professional Learning Teams

In the context of assessment and learning as outlined in the preceding discussion, formative assessment has to be an identification of the appropriate level of development for a scaffolded intervention by the teacher, in order to facilitate a student's progress. It is not the identification of a weakness, lack of skill or error. This is an important point in making maximum use of formative assessment. It is also not true that any form of continuous assessment is formative. Unless they are linked to a developmental progression, practices called 'formative assessment' can be relatively useless and can even be detrimental to student learning.

Formative decisions based on developmental assessment also define the *intervention* role of a teacher. It is a truism that students learning at different levels on a developmental progression need different teaching strategies. Formative assessment in a developmental model focuses on the level of readiness to learn, not the errors, deficits or flaws. Where this is not recognised, it is possible for the teacher to fall into the trap of teaching to a test and, while this might improve a test score, it does not necessarily improve ability or generalised development. (For example, coaching for an intelligence test may improve the IQ score, but not the person's intelligence.) Once a student's level of development is identified, the teacher's decision making shifts from *what* the student needs to learn to *how* the student can best learn at that level (ZPD). This involves the teacher in making decisions about what *intervention strategy* is best for that student at the generalised level of development or *readiness to learn*. When confronted by a range of students at differing levels of development and with differing learning needs and styles, the teacher may need to use a range of teaching strategies even with small groups of students. Practical experience suggests that students can be grouped by levels of development, and a teacher does not have to individualise every aspect of teaching and learning, but classroom management is affected.

Because there will inevitably be a range of possible intervention strategies, just as there are students at different levels of development, *resources* needed for each level also have to be identified, and teachers working alone often need support. Discussion, monitoring and evaluation by the teachers targeting instructional strategies help to clarify and spread the accountability between teachers within schools. There is evidence linking formative use of assessment from standardised tests to the improvement of student learning outcomes through critical and collaborative analysis and discussion of data (for example, Phillips, McNaughton, & MacDonald, 2004; Halverson, Grigg, Prichett, & Thomas, 2005). Evidence-based problem analysis focusing on teacher activities and student achievement is an effective form of professional development that links assessment data directly to teaching, using the evidence for discussion in professional learning teams (PLTs) (Hawley & Valli, 1999). Teachers need to discuss with their colleagues their materials and interventions for each student or group of students.

In addition to monitoring student developmental learning, teachers need a process with which to analyse the data, link them to their own teaching and test the links using evidence in PLTs. The role of these teams is important to the improvement of student learning. Teachers need the opportunity to test their understanding of the data, to propose their strategy and resource allocation and to have their colleagues examine these interpretations and strategies from a critical perspective. When this is done in teams for each learning sequence for the students at different levels on the continuum, real professional development occurs, accountability to peers is inbuilt and teachers get reinforcement from their peers. The students are the eventual winners.

If data are used this way, it is imperative that teachers understand their own practice and how it relates to student achievement. Critical and collaborative discussions where teachers test their theories about these links in PLTs are an important vehicle for doing just that. Discussion and analysis as a component of professional development have been shown to improve teaching and student achievement (Timperley & Robinson, 2001) and are an effective form of professional development in comparison to traditional workshop models (Hawley & Valli, 1999). Worthwhile and significant change in teaching practice can occur when teachers are engaged in examining their own theories of practice (Bransford, Derry, Berliner, & Hammermass, 2005; Hawley & Valli, 1999; Richardson, 1990).

It is of course important for teachers to reflect critically on their own practice individually, but doing so collaboratively has been linked to improved student achievement (Ladson-Billings & Gomez, 2001; Phillips et al., 2004) and changed teacher perceptions (Timperley & Robinson, 2001). Collaborations in PLTs enable teachers to have access to a greater number and divergence of theories against which to test their theories, particularly if the community draws on differing expertise, but it can be a slow and painful process (Ladson-Billings & Gomez, 2001). It does, however, instil a peer approach to accountability within the team and enables each teacher to draw on the expertise and experience of their colleagues. Learning teams of teachers and school leaders, policy makers and researchers can accelerate learning, but the collaborations are only effective if they involve rigorous examinations of teaching and learning, rather than comfortable collaborations in which ideas are simply shared (Ball & Cohen, 1999; Robinson & Lai, 2006).

Having and using the assessment tools, however, are insufficient conditions for teachers to inform their teaching (Halverson et al., 2005). Using standardised assessments formatively also requires the tests to have sufficient diagnostic capacity for teachers to monitor students' learning developmentally. Teachers need to be able to access and interpret assessment data at both the group and individual levels. The diagnostic, clinical or formative capacity of the test must be linked to the identification of the latent variable (the developmental progression). Test scores and item-level (right/wrong) data can be detrimental to formative use of tests. Unless the test items act as a cohesive set and allow the levels of clusters of items to identify underpinning skills, the diagnostic information is minimised and perhaps even counterproductive. When the items act as a cohesive set, they allow an underlying variable to be recognised and used in a formative process.

Different interventions need different *resources* to be identified, acquired, used and evaluated. It may be that the same resource can be used for different levels of development; it may be that the same level of development and the same skill acquisition needs different resources for different students with different learning needs. Matching resources and intervention strategies to student readiness and learning styles is a complex professional skill and one that needs well-developed classroom management skills and resource use. They are professional decisions that teachers make when faced with students at different developmental levels, and with differing developmental needs and differing learning styles. Discussion among teachers can and does help to clarify these decisions and the teachers gain critical support from their colleagues. The importance of sharing and discussing these strategies and resources as part of a learning team is an important step forward.

## Can It Be Done? Reading Comprehension Through Collaborative Learning Teams

### *The Literacy Assessment Project*

In 2004, the Catholic Education Office of Melbourne (CEOM) trialled a range of reading comprehension assessment instruments in 20 Catholic primary schools to examine the benefits and limitations of each. The project evolved in response to schools seeking advice on what was an appropriate approach to the assessment of reading comprehension in years 3 and 4. A whole-of-school commitment to, and by PLTs, was required in order for schools to be accepted into the project.

#### Assessments Administered

The project involved the use of standardised tests administered to students in years 3 and 4. At the beginning and end of an 8-month period, teachers were asked to administer two reading assessments: a year 3 AIM Reading test (VCAA, 2003–2005) and one of a TORCH (Mossensen, Hill, & Masters, 1993), PROBE (Pool, Parkin, & Parkin, 2002) or DART test (ACER, 1994). It represented considerable work for the teachers who had to mark the tests and record for every student whether the answer to each item was correct. A total of 70 teachers administered the tests to approximately 1640 students each year. The assessment data were analysed to develop a *Progression of Reading Development*, discussed next.

#### Establishing the Progression of Reading Development

Common item-equating procedures were used to map all the items from the four tests onto the same continuum (Write & Stone, 1979). All but four of the 236 test questions mapped onto a single underlying variable. The skills audit was then used to interpret the variable and a common *Progression of Reading Development* of eight levels was identified. Students were placed at one of the levels according to their test performance.

The test items were content-analysed and calibrated using item-response theory (Rasch, 1960) to determine whether they provided similar information and whether they could be used interchangeably to report progress and identify starting points for learning. As mentioned, a skills audit was performed. The TORCH, DART and

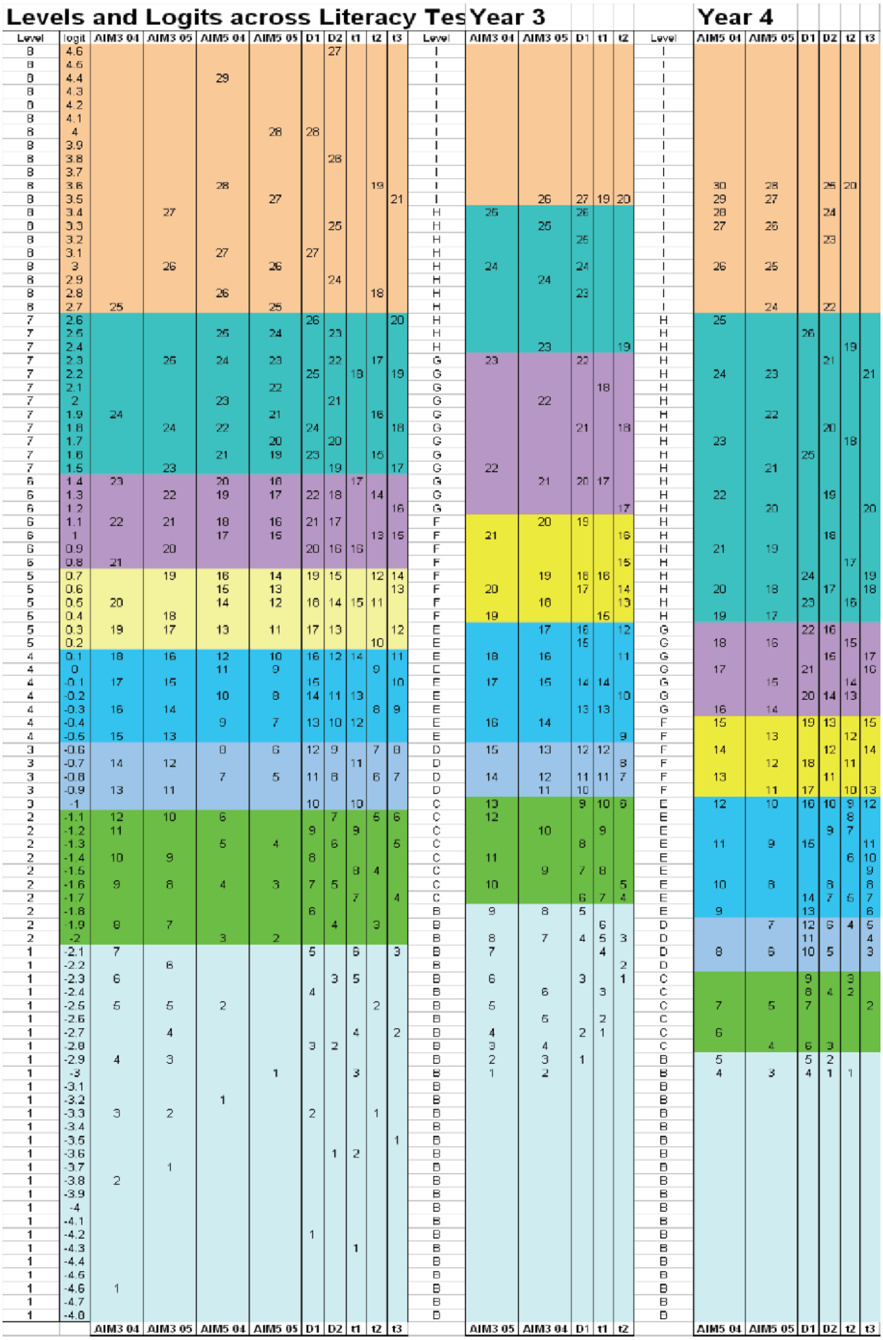
AIM Reading test developers provided some skill descriptors as part of the test documentation, but the PROBE test required a complete item skills audit. The PROBE test questions were classified by the test authors according to six elements of comprehension: literal comprehension, reorganisation, inference, vocabulary, evaluative comprehension and reaction. In some cases, it was difficult to distinguish between the elements of PROBE and confirm the specific skill required to complete a question. When all test items were audited, the skill descriptors were arranged in order of item difficulty and this gave a series of ordered descriptions from which a developmental reading progression emerged. The progression had eight levels. Students were placed at one of the levels according to their two test scores, but a very small number of students had inconsistent scores across their two tests. One test might have had a high score but the second a low score. There were not many of these, but they posed a problem for interpretation of student performance. A decision was taken to use the higher of the two scores.

The teachers were provided with a look-up table that enabled a conversion from a raw score to a level (see Fig. 10.3). Item-response logit values also showed that the width of each of the levels 1–8 were 2, 1.1, 0.4, 0.8, 0.6, 1.7, 1.2 and 2 logits, respectively. The widths of the band levels were important to understand the magnitudes of the growth patterns demonstrated by the students. Repeated studies of development have shown that a growth of 0.5 logits can be expected with a 1-year program for the typical student. This developmental progression therefore represents better than 8 years of development in reading comprehension. The look-up table contains information from the AIM tests, the TORCH and the DART tests. No data have been included in the table for the PROBE because the subtests in the PROBE package were unreliable, and it seriously affected the vertical equating procedures.

The data were collected at the beginning and end of the first year and again in the second year. It became clear that there were irregularities with data from students taking the PROBE test. This was traced to the low reliability of the PROBE test, which was attributed to the design of the tests in the PROBE package. A short, seven-item test, for example, mapped onto an eight-level scale led to unstable results. Compare this to a 45-item test as in the AIM, mapped to eight levels. The problem was accentuated by teacher judgment in scoring the PROBE because of inconsistency in marking stringency. However, teachers indicated that PROBE was a useful tool because it provided exercises in each of the cognitive classifications and the teachers valued the advice on teaching strategies and resources to meet individual students' learning needs. So, while PROBE went some way to providing this advice, the DART, TORCH and AIM Reading tests had greater stability as assessments for progress reporting.

## The Professional Learning Teams

Figure 10.4 shows the three collaborating organisations involved in the project. The CEOM oversaw the project. Teachers from years 3 and 4 from each school



| Levels and Logits across Literacy Tes | | | | | | | | | | | | Year 3 | | | | | | Year 4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Level | logit | AIM3 04 | AIM3 05 | AIM5 04 | AIM5 05 | D1 | D2 | t1 | t2 | t3 | Level | AIM3 04 | AIM3 05 | D1 | t1 | t2 | Level | AIM5 04 | AIM5 05 | D1 | D2 | t2 | t3 |
| 8 | 4.6 | | | | | | 27 | | | | I | | | | | | I | | | | | | |
| 8 | 4.5 | | | | | | | | | | I | | | | | | I | | | | | | |
| 8 | 4.4 | | | 29 | | | | | | | I | | | | | | I | | | | | | |
| 8 | 4.3 | | | | | | | | | | I | | | | | | I | | | | | | |
| 8 | 4.2 | | | | | | | | | | I | | | | | | I | | | | | | |
| 8 | 4.1 | | | | | | | | | | I | | | | | | I | | | | | | |
| 8 | 4 | | | | 28 | 28 | | | | | I | | | | | | I | | | | | | |
| 8 | 3.9 | | | | | | | | | | I | | | | | | I | | | | | | |
| 8 | 3.8 | | | | | | 26 | | | | I | | | | | | I | | | | | | |
| 8 | 3.7 | | | | | | | | | | I | | | | | | I | | | | | | |
| 8 | 3.6 | | | 28 | | | | | 19 | | I | | | | | | I | 30 | 28 | | 25 | 20 | |
| 8 | 3.5 | | | | 27 | | | | | 21 | I | | 26 | 27 | 19 | 20 | I | 29 | 27 | | | | |
| 8 | 3.4 | | 27 | | | | | | | | H | 26 | | 26 | | | I | 28 | | | 24 | | |
| 8 | 3.3 | | | | | | 25 | | | | H | | 25 | | | | I | 27 | 26 | | | | |
| 8 | 3.2 | | | | | | | | | | H | | | 25 | | | I | | | | 23 | | |
| 8 | 3.1 | | | 27 | | 27 | | | | | H | | | | | | I | | | | | | |
| 8 | 3 | | 26 | | 26 | | | | | | H | 24 | | 24 | | | I | 26 | 25 | | | | |
| 8 | 2.9 | | | | | | 24 | | | | H | | 24 | | | | I | | | | | | |
| 8 | 2.8 | | | 26 | | | | | 18 | | H | | | 23 | | | I | | | | | | |
| 8 | 2.7 | 25 | | | 25 | | | | | | H | | | | | | I | | 24 | | 22 | | |
| 7 | 2.6 | | | | | 26 | | | | 20 | H | | | | | | H | 25 | | | | | |
| 7 | 2.5 | | | 25 | 24 | | 23 | | | | H | | | | | | H | | | 26 | | | |
| 7 | 2.4 | | | | | | | | | | H | | 23 | | | 19 | H | | | | | 19 | |
| 7 | 2.3 | | 25 | 24 | 23 | | 22 | | 17 | | G | 23 | | 22 | | | H | | | | 21 | | |
| 7 | 2.2 | | | | | 25 | | 18 | | 19 | G | | | | | | H | 24 | 23 | | | | 21 |
| 7 | 2.1 | | | | 22 | | | | | | G | | | | 18 | | H | | | | | | |
| 7 | 2 | | | 23 | | | 21 | | | | G | | 22 | | | | H | | | | | | |
| 7 | 1.9 | 24 | | | 21 | | | | 16 | | G | | | | | | H | | 22 | | | | |
| 7 | 1.8 | | 24 | 22 | | 24 | | | | 18 | G | | | 21 | | 18 | H | | | | 20 | | |
| 7 | 1.7 | | | | 20 | | 20 | | | | G | | | | | | H | 23 | | | | 18 | |
| 7 | 1.6 | | | 21 | 19 | 23 | | | 15 | | G | | | | | | H | | | 25 | | | |
| 7 | 1.5 | | 23 | | | | 19 | | | 17 | G | 22 | | | | | H | | 21 | | | | |
| 6 | 1.4 | 23 | | 20 | 18 | | | 17 | | | G | | 21 | 20 | 17 | | H | | | | | | |
| 6 | 1.3 | | 22 | 19 | 17 | 22 | 18 | | 14 | | G | | | | | | H | 22 | | | 19 | | |
| 6 | 1.2 | | | | | | | | | 16 | G | | | | | 17 | H | | 20 | | | | 20 |
| 6 | 1.1 | 22 | 21 | 18 | 16 | 21 | 17 | | | | F | | 20 | 19 | | | H | | | | | | |
| 6 | 1 | | | 17 | 15 | | | | 13 | 15 | F | 21 | | | | 16 | H | | | | 18 | | |
| 6 | 0.9 | | 20 | | | 20 | 16 | 16 | | | F | | | | | | H | 21 | 19 | | | | |
| 6 | 0.8 | 21 | | | | | | | | | F | | | | | 15 | H | | | | | 17 | |
| 5 | 0.7 | | 19 | 16 | 14 | 19 | 15 | | 12 | 14 | F | | 19 | 18 | 16 | | H | | | 24 | | | 19 |
| 5 | 0.6 | | | 15 | 13 | | | | | 13 | F | 20 | | 17 | | 14 | H | 20 | 18 | | 17 | | 18 |
| 5 | 0.5 | 20 | | 14 | 12 | 18 | 14 | 15 | 11 | | F | | 18 | | | 13 | H | | | 23 | | 16 | |
| 5 | 0.4 | | 18 | | | | | | | | F | 19 | | | 16 | | H | 19 | 17 | | | | |
| 5 | 0.3 | 19 | 17 | 13 | 11 | 17 | 13 | | | 12 | E | | 17 | 16 | | 12 | G | | | 22 | 16 | | |
| 5 | 0.2 | | | | | | | | 10 | | E | | | 15 | | | G | 18 | 16 | | | 15 | |
| 4 | 0.1 | 18 | 16 | 12 | 10 | 16 | 12 | 14 | | 11 | E | 18 | 16 | | | 11 | G | | | | 16 | | 17 |
| 4 | 0 | | | 11 | 9 | | | | 9 | | E | | | | | | G | 17 | | 21 | | | 16 |
| 4 | -0.1 | 17 | 15 | | | 15 | | | | 10 | E | 17 | 15 | 14 | 14 | | G | | 15 | | | 14 | |
| 4 | -0.2 | | | 10 | 8 | 14 | 11 | 13 | | | E | | | | | 10 | G | | | 20 | 14 | 13 | |
| 4 | -0.3 | 16 | 14 | | | | | | 8 | 9 | E | | | 13 | 13 | | G | 16 | 14 | | | | |
| 4 | -0.4 | | | 9 | 7 | 13 | 10 | 12 | | | E | 16 | 14 | | | | F | 15 | | 19 | 13 | | 15 |
| 4 | -0.5 | 15 | 13 | | | | | | | | E | | | | | 9 | F | | 13 | | | 12 | |
| 3 | -0.6 | | | 8 | 6 | 12 | 9 | | 7 | 8 | D | 15 | 13 | 12 | 12 | | F | 14 | | | 12 | | 14 |
| 3 | -0.7 | 14 | 12 | | | | | 11 | | | D | | | | | 8 | F | | 12 | 18 | | 11 | |
| 3 | -0.8 | | | 7 | 5 | 11 | 8 | | 6 | 7 | D | 14 | 12 | 11 | 11 | 7 | F | 13 | | | 11 | | |
| 3 | -0.9 | 13 | 11 | | | | | | | | D | | 11 | 10 | | | F | | 11 | 17 | | 10 | 13 |
| 3 | -1 | | | | | 10 | | 10 | | | C | 13 | | 9 | 10 | 6 | E | 12 | 10 | 16 | 10 | 9 | 12 |
| 2 | -1.1 | 12 | 10 | 6 | | | 7 | | 5 | 6 | C | 12 | | | | | E | | | | | 8 | |
| 2 | -1.2 | 11 | | | | 9 | | 9 | | | C | | 10 | | 9 | | E | | | | 9 | 7 | |
| 2 | -1.3 | | | 5 | 4 | | 6 | | | 5 | C | | | 8 | | | E | 11 | 9 | 15 | | | 11 |
| 2 | -1.4 | 10 | 9 | | | 8 | | | | | C | 11 | | | | | E | | | | | 6 | 10 |
| 2 | -1.5 | | | | | | | 8 | 4 | | C | | 9 | 7 | 8 | | E | | | | | | 9 |
| 2 | -1.6 | 9 | 8 | 4 | 3 | 7 | 5 | | | | C | 10 | | | | 5 | E | 10 | 8 | | 8 | | 8 |
| 2 | -1.7 | | | | | | | 7 | | 4 | C | | | 6 | 7 | 4 | E | | | 14 | 7 | 6 | 7 |
| 2 | -1.8 | | | | | 6 | | | | | B | 9 | 8 | 5 | | | E | 9 | | 13 | | | 6 |
| 2 | -1.9 | 8 | 7 | | | | 4 | | 3 | | B | | | | 6 | | D | | 7 | 12 | 6 | 4 | 5 |
| 2 | -2 | | | 3 | 2 | | | | | | B | 8 | 7 | 4 | 5 | 3 | D | | | 11 | | | 4 |
| 1 | -2.1 | 7 | | | | 5 | | 6 | | 3 | B | 7 | | | 4 | | D | 8 | 6 | 10 | 5 | | 3 |
| 1 | -2.2 | | 6 | | | | | | | | B | | | | | 2 | D | | | | | | |
| 1 | -2.3 | 6 | | | | | 3 | 5 | | | B | 6 | | 3 | | 1 | C | | | 9 | | 3 | |
| 1 | -2.4 | | | | | 4 | | | | | B | | 6 | | 3 | | C | | | 8 | 4 | 2 | |
| 1 | -2.5 | 5 | 5 | 2 | | | | | 2 | | B | 5 | | | | | C | 7 | 5 | 7 | | | 2 |
| 1 | -2.6 | | | | | | | | | | B | | 6 | | 2 | | C | | | | | | |
| 1 | -2.7 | | 4 | | | | | 4 | | 2 | B | 4 | | 2 | 1 | | C | 6 | | | | | |
| 1 | -2.8 | | | | | 3 | 2 | | | | B | 3 | 4 | | | | C | | 4 | 6 | 3 | | |
| 1 | -2.9 | 4 | 3 | | | | | | | | B | 2 | 3 | 1 | | | B | 5 | | 5 | 2 | | |
| 1 | -3 | | | | 1 | | | 3 | | | B | 1 | 2 | | | | B | 4 | 3 | 4 | 1 | 1 | |
| 1 | -3.1 | | | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -3.2 | | | 1 | | | | | | | B | | | | | | B | | | | | | |
| 1 | -3.3 | 3 | 2 | | | 2 | | | 1 | | B | | | | | | B | | | | | | |
| 1 | -3.4 | | | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -3.5 | | | | | | | | | 1 | B | | | | | | B | | | | | | |
| 1 | -3.6 | | | | | | 1 | 2 | | | B | | | | | | B | | | | | | |
| 1 | -3.7 | | 1 | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -3.8 | 2 | | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -3.9 | | | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -4 | | | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -4.1 | | | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -4.2 | | | | | 1 | | | | | B | | | | | | B | | | | | | |
| 1 | -4.3 | | | | | | | 1 | | | B | | | | | | B | | | | | | |
| 1 | -4.4 | | | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -4.5 | | | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -4.6 | 1 | | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -4.7 | | | | | | | | | | B | | | | | | B | | | | | | |
| 1 | -4.8 | | | | | | | | | | B | | | | | | B | | | | | | |
| | | AIM3 04 | AIM3 05 | AIM5 04 | AIM5 05 | D1 | D2 | t1 | t2 | t3 | | AIM3 05 | AIM3 04 | D1 | t1 | t2 | | AIM5 04 | AIM5 05 | D1 | D2 | t2 | t3 |

**Fig. 10.3** Look-up table for teachers to convert raw scores to levels

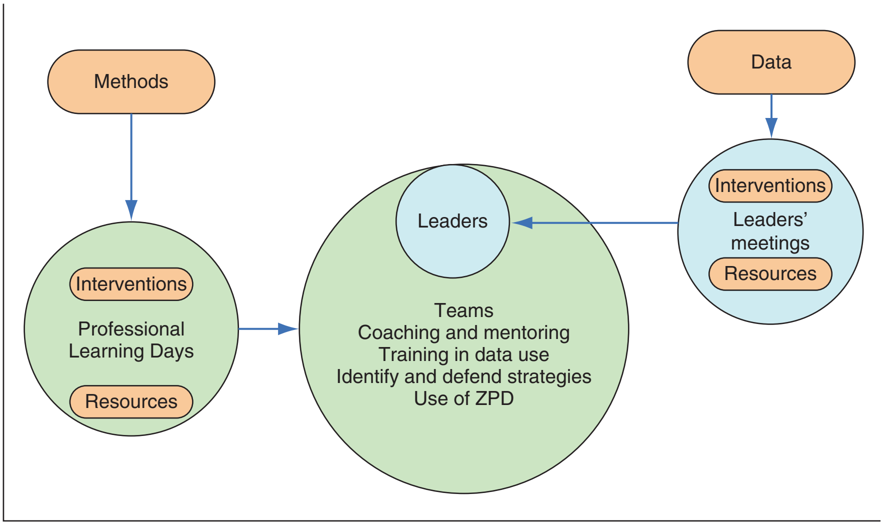


**Fig. 10.4** Professional learning teams in a collaborative partnership project

formed the PLTs. The PLT leadership role[1] was undertaken by the school's literacy coordinator. The university provided inputs in two areas. The first was in the field of assessment and the use of data in a developmental assessment framework. The second was in providing an overview of reading comprehension and the associated intervention strategies.

PLTs worked to investigate and extend their knowledge of strategies for comprehension and began to identify a range of intervention strategies that could assist the development of reading skills for students at different comprehension levels. The literacy team leaders also attended professional learning sessions to reflect on the data and build a common understanding of its interpretation and links to teaching strategies. This was consistent with Joyce and Showers (2002), who argued that unless structured opportunities for professional development and training were provided, teachers found it difficult to acquire new teaching skills; and unless adequate follow-up and support was provided in the school, the new skills do not transfer into the classroom. Teachers needed to control their professional learning as a key to the improvement in the quality of student learning and teacher effectiveness.

The PLTs operated in a tiered approach. Each school-based literacy team was led by the coordinator, who in turn belonged to a leaders' team. The leaders' team met with the project specialists and shared results, discussed and critiqued the assessments. They were, at a school level, accountable to other team leaders for their developing expertise and the way in which they understood and used the data. They

[1] The leadership role was critical as the PLT leader had to be involved in all project meetings. The PLT leader was also given the opportunity to be both a lead learner who had the 'big picture' of the project and as such could contribute to overall the project design and management.

were also responsible for explaining the evidence to their literacy teachers in the school-based team and for reporting to the leaders' team and the project management the impact and effect on the literacy PLTs. The tiered accountability model ensured that external expertise could be distributed across the schools in an efficient manner. It also meant that accountability was operating at two levels. Team leaders were accountable to other team leaders for the way in which they understood, used and explained the evidence to their colleagues in the PLTs and how this led to improvements in student learning at their schools. Teachers were accountable to other teachers in their school for the way in which the evidence of development was used to identify intervention strategies and relevant resources. They needed to link the developmental level of individual students to an intervention strategy and then discuss it with their colleagues. Team leaders had to summarise the overall strategies of the literacy team and link this to the development and evidence of growth that the data showed for their school. The external specialists were accountable for the quality of the data and their reporting materials that the team leaders and the literacy teachers were using in the schools. It was a multi-tiered accountability approach for the use of data in intervention, and student growth and development in reading comprehension.

School-level aggregated data always showed growth, but this was not surprising given the effects of maturation and normal expected progress for each year level. An example of a typical school's data is shown in Fig. 10.5. The horizontal axis represents the eight levels in the reading comprehension scale. The vertical axis represents the percentage of students at each level. The two superimposed bar charts represent the assessments at October in each of 2005 and 2006. The shift to the right is interpreted as growth. Was it just maturation? It might be, but if it were, it was astonishing and uniform, large, maturation effects across the 20 schools. Based on large-scale studies using item-response analyses, there is evidence of a substantial shift in reading comprehension development. The difference in the item-response measure (logits) between the high and low levels is more than 8 logits. This was

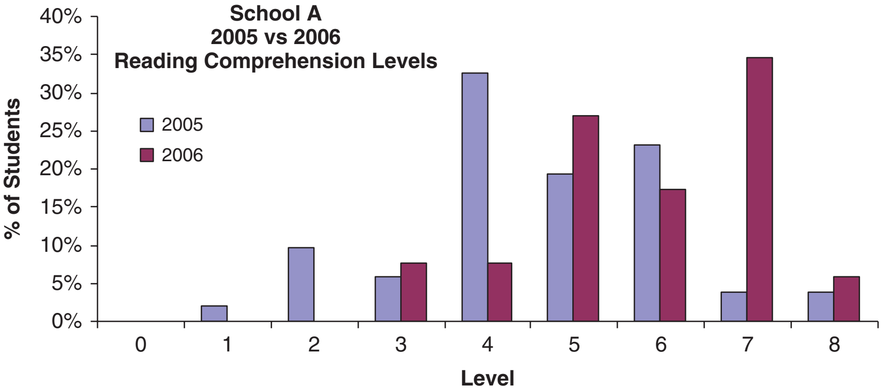


**Fig. 10.5** Gains from 2005 to 2006 by years 3 and 4 for school A

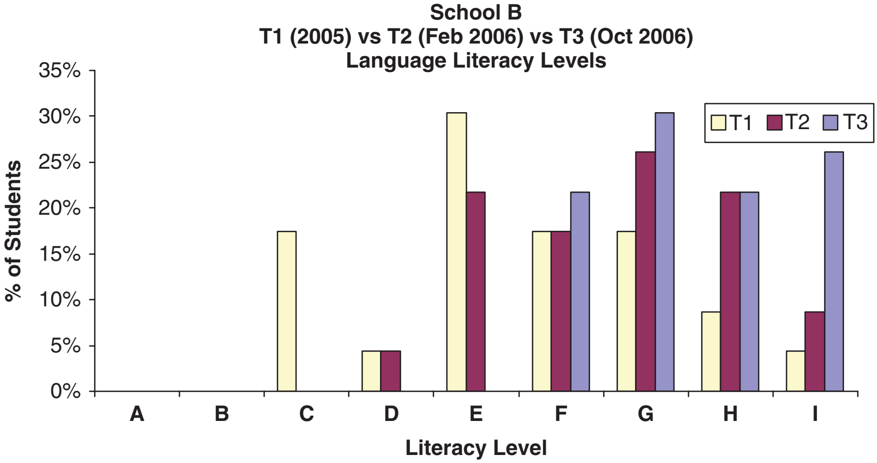


**Fig. 10.6** Three testing periods, October 2005, March/April 2006 and October 2006 for school B

approximately 1 logit per reading comprehension level (see Fig. 10.6). As noted, a look-up table was provided to teachers for translating test scores into logits and levels of development. In national, state and international studies the general gain per year level is 0.5 logits per school year. This includes general gains on the AIM test used in this project when it is used in a state-level cohort testing. The average gain per school in this study was approximately 1–1.5 levels or 1.5 logits—three times the normal expected gain. If this is maturation, it is extraordinary. However, the average gain is not the only way to describe the shift. Consider the lowest group. They have moved upwards by two levels and one school had improved by 2.7 logits—four to more than five times the expected growth! Less growth is evident at the upper levels, but this could be because of the limits of the measurement instruments. In any pre-test and post-test design, as used in this program, students who are lower initial performers will appear to grow or improve more than those students at higher initial performance levels. This effect, known as 'regression-to-the-mean', will occur regardless of any intervention (Campbell & Russo, 1999). So while the gains in the lower levels are impressive, some might be attributed to maturation, some to regression and some to practice effect due to the retesting procedures. However, gains are still up to five times the expected gain; gains of such magnitude cannot be dismissed as attributable to design threats to validity.

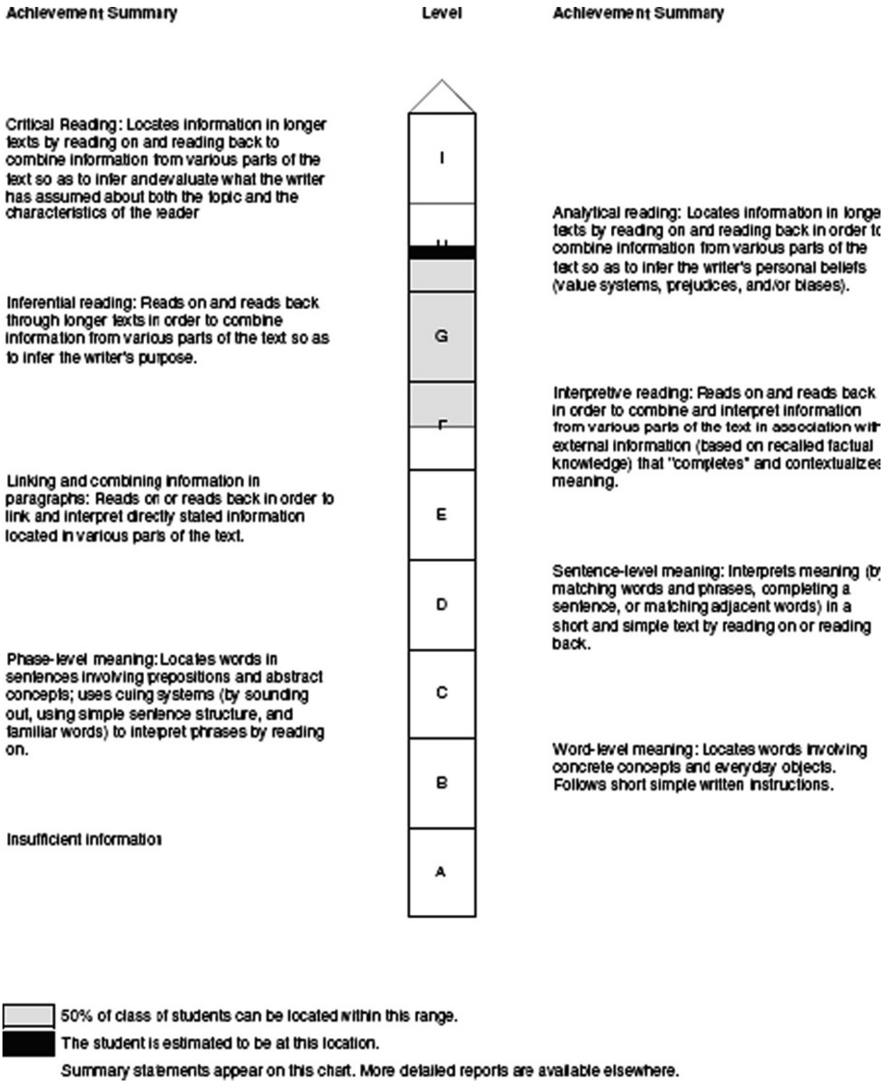
The collaborative basis of the three-tiered PLTs recognised and developed the knowledge, expertise and skills that the project team brought to it. The results of the *Progression of Reading Development* were provided to the teachers in a discussion forum. Professional learning was shared in these forums and both literacy coordinators and project team members gained insights from each other as the discussions cantered on the application of the *Progression of Reading Development* and its application to targeted and differentiated teaching (Griffin, 2007; Perkins, 2005). Close liaison was maintained between the university, the CEOM and the literacy coordinators in each of the schools. When the literacy leaders examined the data with the CEOM and university staff, no attempts were made to identify the appropriate teaching strategies. The main point was the recognition that intervention was needed, and targeted intervention was essential for students at different levels of development. The emphasis was always on *what* had to be learned and the team leaders then took this information to the professional learning teams in the school to work on strategies and *how* learning was to take place.

The work of *team leaders* was also central to the process of developing the teachers' confidence in using an evidence-based developmental framework. All members of the PLT examined the data in light of their knowledge of individual students, and this also assisted in identifying and remedying any anomalies in the data. The team leaders worked with their *teams* to trial and document appropriate teaching strategies and resources. The substantial growth can be seen in the Figs. 10.5 and 10.6 for two schools over two and three assessment periods, respectively.

A series of reports was provided to teachers. The first represented the individual student performance on the developmental continuum. This has been labelled the 'rocket' report (see Fig. 10.7), and it presents a mix of criterion-referenced and norm-referenced interpretations of the student performance data. The mark on the spine of the report indicates the student's position on the developmental continuum. The box represents the inter-quartile range and enables an interpretation of the student's performance relative to the other students in the school. The descriptions at the side of the 'rocket' describe a summary of each developmental level. The level adjacent to the black marker is a summary description of the ZPD where scaffolding can best be used. It is at this level of development that teachers were encouraged to identify intervention strategies.

The second set of data was the class report (see Fig. 10.8). In this report, it was possible to see, for each student, the results of two or three assessments. In Fig. 10.8, the results for October 2005, March and October 2006 are represented by different shades in the columns in the chart. The descriptions across the top of the report are identical to those in the 'rocket' chart. The top of the bar for each student is at the same position as the marker in the 'rocket'. It indicates the level at which the student is developing. The shaded region is the inter-quartile region for the most recent assessment. The report shows how much progress each individual has made. The report also shows how the rate of change is relative to the group rate of change. The teacher is given an overall perspective of the class and individual achievement levels, rate of change of achievement and the effect of intervention for each individual. It also helps to identify relatively homogeneous intervention groups. It is clear that not all students progress equally, or even at all. Some regress. This was difficult for teachers to accept and, at times, predictably from a teachers' point of view, cast doubt on the measures, but working through the data for each student item by item soon showed that the data were accurate, and the student performance was erratic. Measurement error appeared to have been influenced by student engagement in the assessment as well as by instrument effects.

Team-debriefing sessions at school used these presentations of data analyses by the team leaders, who were trained in the use of the reporting software. The



Critical Reading: Locates information in longer texts by reading on and reading back to combine information from various parts of the text so as to infer andevaluate what the writer has assumed about both the topic and the characteristics of the reader

Analytical reading: Locates information in longe texts by reading on and reading back in order tc combine information from various parts of the text so as to infer the writer's personal beliefs (value systems, prejudices, and/or biases).

Inferential reading: Reads on and reads back through longer texts in order to combine information from various parts of the text so as to infer the writer's purpose.

Interpretive reading: Reads on and reads back in order to combine and interpret information from various parts of the text in association wit external information (based on recalled factual knowledge) that "completes" and contextualize meaning.

Linking and combining information in paragraphs: Reads on or reads back in order to link and interpret directly stated information located in various parts of the text.

Sentence-level meaning: Interprets meaning (b matching words and phrases, completing a sentence, or matching adjacent words) in a short and simple text by reading on or reading back.

Phase-level meaning: Locates words in sentences involving prepositions and abstract concepts; uses cuing systems (by sounding out, using simple sentence structure, and familiar words) to intepret phrases by reading on.

Word-level meaning: Locates words involving concrete concepts and everyday objects. Follows short simple written instructions.

Insufficient informatio

▭ 50% of class of students can be located within this range.

■ The student is estimated to be at this location.

Summary statements appear on this chart. More detailed reports are available elsewhere.

**Fig. 10.7** Criterion report for an individual student

teams discussed modifications in teaching, differentiated intervention and targeting in teaching strategies. Project team members held separate meetings with the team leaders. After 1 year, there had been a substantial change in discourse, intervention practice and resource use linked to change in student literacy development. School teams also selected an area of inquiry about the learning and teaching of reading and their investigation was linked to both student and teacher learning. It helped to highlight the importance of the assessment data. At project meetings, team leaders shared with colleagues from other schools their resources for teaching intervention, and these materials were prepared for a website for all schools to use.

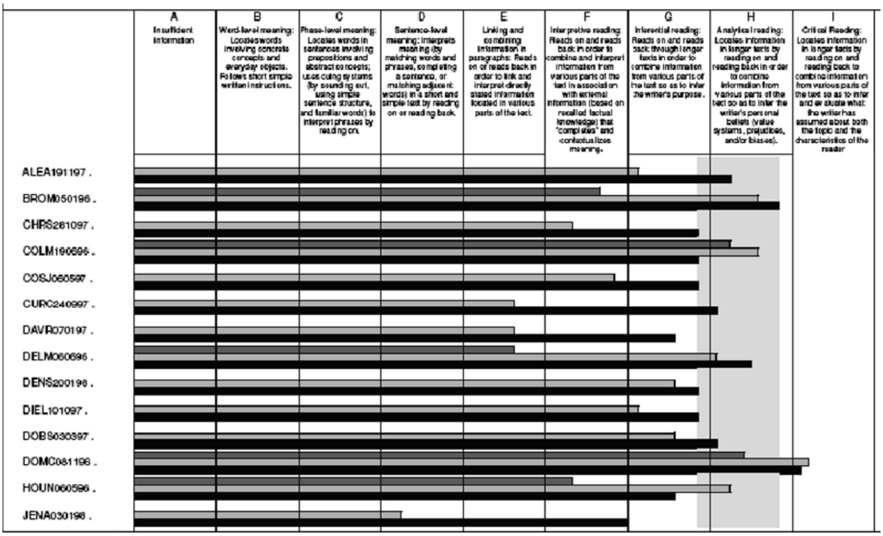


**Fig. 10.8** Class-level report showing change against levels for individual students

The importance of PLTs cannot be overstated. Each school allocated time for their teams to meet and examine the data and its connection to their intervention practices. The team leaders shared teaching experiences among colleagues during team meeting days and with leaders from other schools during project meetings. Through these professional learning sessions, all teachers had the opportunity to engage with new and challenging understandings that were directly relevant to improving student outcomes in reading. Professional learning opportunities were drawn from research, ensuring that input was theoretically based and situated within school data. The knowledge and experience bases of teachers were valued and incorporated into the theoretical framework upon which the work was based. Consideration was given to ensuring there was a mix of both input from outside experts and opportunities for teachers to work through issues and engage in learning activities. The ongoing nature of the project, with a consistent cohort of schools and team leaders engaged in the project for over 4 years, has provided time for an action research cycle to occur, with an emphasis on reflective practice (Kemmis & McTaggart, 2000).

## *Project Outcomes*

An analysis of student data over the period of 2005 and 2006 indicated that students had made progress as measured across the developmental progression. Not only had the cohort moved up the scale, but also the spread had not increased. This suggested that all students were developing and the 'tail' of the distribution was not being left behind or remaining static, as was the expected case if the data from Rowe and Hill (1996) study were to be replicated. It was also clear from the measures described above that a year's expected gain (about 0.5 logit) was exceeded many times over by groups, but there were also individual students who appeared to regress. Teachers set about specific intervention with those individuals but always emphasised the readiness issue in determining the 'what' and 'how' of teaching and learning.

The project's success in 'raising the bar' and 'closing the gap' (Fullan, 2005) coincided with a deepening of teacher knowledge and discourse. Confronting the teachers with the evidence of student development and the heterogeneity of the class with respect to reading comprehension development caused an inescapable shift in emphasis.

At no stage were any teaching strategies prescribed, or even recommended. Teachers were shown the data, provided with an interpretation and asked what they would do about it. An external opportunity for discussion and discourse on reading instruction was attended by all members of the school teams after the leader had had a chance to discuss class and individual student results with their team members. They had also had a chance to search for suitable intervention strategies and resources before the professional development day.

A marked shift was identified in the discourse of the literacy team leaders. The same was true of the in-school team members due to their engagement in discussions about the targeted and clinical, interventionist evidence-based approaches to learning and teaching of reading. At the beginning of the project, the initial focus of the discourse centred on the reading acquisition as a discrete set of skills to be taught, learned, practiced and applied. It also focused on resources for their own sake without connection to the level of development of the students. Assessments were used to identify students' mastery (or non-mastery) of discrete skills, and intervention was viewed as a direct approach to teaching specific skills that had not been mastered, using texts and other resources written for specific skill acquisition. It was a clear example of a deficit approach to teaching and intervention. The reading curriculum was being defined in terms of discrete skill acquisition. After the first annual cycle of data interpretation and targeted intervention, aimed at personalised learning plans for students at each level on the developmental progression, the discourse had changed, the view of reading development had changed, the approach to intervention had changed and, more importantly, the results showed obvious gains for students. Every school group had moved upwards on the scale, some more than others. The 'tail' was moving up at the same or a better rate than the 'top'.

A developmental progression on its own, however, will not result in student learning improvement unless there is accompanying change in teacher behaviour and a relevant change in curriculum and resources. When a developmental progression was used in conjunction with targeted instruction, gains were achieved. Changes in teacher behaviour depended on being able to use the evidence appropriately and these, in turn, were dependent on opportunities to learn from externally provided professional development at team leader level, internal development within the teams and whole-of-team professional development provided externally. The combination led to whole-of-school changes in pedagogy. It was never assumed that teachers working in isolation had the expertise or opportunity to design the effective learning opportunities to move the students along the continuum, even when they had identified a starting point for a student's learning. This understanding developed with exposure to evidence, experience in PLTs and opportunities to learn abut data and its links to intervention, and their accountability to each other as members of the learning teams.

### *Perceptions of Professional Learning*

In the PLTs, teachers acknowledged that they generally had a range of data that they consistently used. The practice of placing students on the developmental continuum not only confirmed their teacher perceptions but also gave them a framework and a language that enabled their perceptions to be shared with other members of the team. This meant that data were examined by the entire team, and that all members of the team shared teaching and learning needs and addressed teaching strategies, supporting each other. They gained input from both the assessment project data, from the professional team learning days (offsite, in which specialists presented on teaching strategies for reading development) and from their team meetings (onsite).

Identifying students' levels on the developmental reading progression or their ZPD for scaffolding purposes were not on their own translated directly into effective teaching. This decision needed opportunities for the teachers to develop their knowledge of developmental learning and their understanding of appropriate targeted intervention practices. The teachers drew on the examples learned at the professional learning days (offsite) and emphasised the importance of their own and their colleagues' knowledge and experience in identifying appropriate intervention strategies together with the need to develop personalised learning plans for each student based on readiness to learn. In order to achieve this, they had to learn to use the data, link the data to an interpretation of each student's development and then match a teaching and resource strategy to the student's readiness to learn. This was the fundamental link between practice and theory, coupled with focused professional reading and professional learning opportunities. It underscored the importance of data-driven instruction accompanied by an emphasis on teacher learning and professional development.

## Implications for Teacher Education

An important series of questions remain. Can this be applied to pre-service teacher education? Can it be developed into packaged in-service teacher education? Is it possible to establish PLTs consisting of teachers and student teachers, with a team leader? Can the teams be given the opportunity to address specific learning issues in a school, supported by the university-backed 'offsite' and 'in-school' professional development with team leaders steering the professional learning 'onsite'? How can a developmental learning progression underpin each target problem, if, for example, the development is in arts, aesthetics or another discipline which does not normally develop this way? Would this be a successful approach in pre-service teacher education?

It is clear that teachers do not work effectively as solo teachers. Some can, but maximum gains are achieved under the following conditions.

1. Teachers need to learn how to work as members of PLTs.
2. Teachers need skills in interpreting and using data to make decisions about teaching and learning intervention.
3. The teams have to have an approach to peer accountability at a within-school level.
4. Accountability has to be linked to the way teachers use data to make decisions about intervention.
5. Intervention decisions have to be matched with the right resources.
6. Team leaders need to be accountable to other team leaders at a between-school level.
7. The team leaders have to be accountable for the way in which they use the training to inform and lead their colleagues at the within-school level.
8. Project leaders and specialists (if there are any) have to be held accountable for their advice and input to the team leaders.
9. Accountability at every level consists of transparent decision making and collaboration with the members of the learning community.
10. Central to the teacher learning teams is the use of an interpretation framework that links learning and teaching. This is usually in the form of a developmental learning progression. Where this is available, teachers have a common framework to identify intervention points and appropriate teaching strategies for individual students.
11. Discussion of these intervention points and resources appropriate to the intervention is an essential aspect of the professional learning team and peer accountability.

## Glossary

**Assessment** A process of gathering, interpreting and using information about learning. The process of gathering can take many forms, from tests to performances or work samples. The interpretation usually involves some form of measurement or coding and their use leads to decisions about teaching and learning

**Calibration** A process that assesses the accuracy of the modelling process described in the item-response modelling explanation. Calibration establishes the errors of measurement and the accuracy of the modelling process

**Construct and latent construct** A construct is a framework we create in our minds to help us understand our observations. An example is 'intelligence', which does not exist, but psychologists use it to explain different levels of cognitive ability demonstrated by people. 'Latent' means unseen or hidden. Usually constructs are hidden and the term latent could be considered redundant. Constructs are hidden because we observe evidence of the construct and then infer the presence or amount of the construct present in a person

**Content analysis** A systematic, qualitative process used to identify common meanings among different verbal descriptions

**Criterion referenced** Criterion-referenced interpretation of performance data focuses only on what a person can do. There is no allowance given for group membership or personal characteristics. In teaching and learning, this is 'what a person can do'; it is not about how they learn

**Developmental progression and developmental continuum** The developmental continuum or progression is a series of descriptions that demonstrate growth in a specific direction; the progressions describe an accumulation of skills and knowledge and can be divided into levels or bands. Glaser called them 'stages', but this might be misinterpreted as a lock-step development

**Item-response modelling** A mathematical procedure that examines how a mathematical equation can be used to describe how people answer questions on a test, questionnaire or observation report. The extent to which the equation can 'fit' the person's responses to test questions is called 'modelling the response patterns'

**Measurement error** Every instrument has error associated with it. Measurement error indicates how accurate a test is in providing information about the amount of a person's cognitive skill. Large errors make the test interpretation invalid and the measurement error is usually reported in terms of a reliability index. Values near zero indicate a poor test. Values near 1.0 indicate an accurate test but do not imply correct interpretation

**Measurement** A process of assigning numbers to things. In education it is a matter of using numerical codes to designate learning. The measures are always codes and measurement needs to provide a way to decode or interpret what the numbers mean

**Psychometric basis** The term psychometric basis of an interpretation means that the issue is considered only in terms of the quantifiable data. The link to learning or teaching implications is not paramount. Psychometrics is an exact science, mathematically based, and needs to be interpreted carefully to decode the information for teaching and learning implications

**Standardised test** A test that is administered in a standardised way. No allowance is made for varying the method of administering the test

**Test instrument** A test. It is common for tests, questionnaires and observation schedules to be called instruments and the questions on them to be called items; hence the test instrument consists of test items to which pupils respond

**Variable** A way of describing how people differ in some specified measures

## References

ACER (Australian Council for Educational Research). (1994). *Developmental assessment resource for teachers*. Melbourne: Author.

Adams, R. J., & Khoo, S. T. (1995). *Quest: Interactive test analysis*. Melbourne: ACER.

Ball, D., & Cohen, D. (1999). Developing practice, developing practitioners: Toward a practice based theory of professional education. In G. Sykes, & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco: Jossey Bass.

Bransford, J., Derry, S., Berliner, D., & Hammermass, K. (2005). Theories of learning and their role in teaching. In L. Darling-Hammond, & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 40–87). San Francisco: John Wiley.

Campbell, D. T., & Russo, M. J. (1999). *Social experimentation*. California: Sage Publications.

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum and Associates.

Fullan, M. (2005). *Leadership and sustainability: System thinkers in action*. California: Corwin Press.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, *18*(5), 19–521.

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist*, *36*, 923–936.

Glaser, R. (1990). Expertise. In M. W. Evnsenk, A. N. Ellis, E. Hunt, & P. Johnson-Laird (Eds.), *The Blackwell dictionary of cognitive psychology.* Oxford, UK: Blackwell Reference.

Greaney, V. S. R., Khandker, S. R., & Alam, M. (1990). *Bangladesh: Assessing basic learning skills.* Bangladesk Development Series. Dhaka: University Press Ltd.

Griffin, P. (1997). *Assessment in schools and workplace*. Inaugural professorial lecture, University of Melbourne, September.

Griffin, P. (1998). *Vietnamese national study of student achievement in mathematics and Vietnamese*. Hanoi: National Institute for Education and Science.

Griffin, P. (2007). The comfort of competence and the uncertainty of assessment. *Studies in Educational Evaluation*, *33*(1), 87–99.

Griffin, P., & Forwood, A. (1990). *The adult literacy and numeracy scales*. Canberra: Department of Education, Employment, Training and Youth Affairs.

Griffin, P., & Nix, P. (1990). *Assessment and reporting: A new approach*. Sydney: Harcourt, Brace, Jovanovic.

Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2005). *The new instructional leadership: Creating data-driven instructional systems in schools* (WCER Working Paper No. 2005–9). Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research. Retrieved November 3, 2005, from <www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2005_9.pdf.

Hawley, W. D., & Valli, L. (1999). The essentials of effective professional development: A new consensus. In L. Darling-Hammond, & G. Sykes (Eds.), *Teaching as a learning profession* (pp. 127–150). San Francisco: Jossey-Bass.

Joyce, B., & Showers, B. (2002). Student achievement through staff development. In B. Joyce (Ed.), *Designing training and peer coaching: Our needs for learning.* VA, USA: ASCD.

Kemmis, S., & McTaggart, R. (2000). Participatory action research. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 567–606). Thousand Oaks: Sage Publications.

Ladson-Billings, G., & Gomez, M. L. (2001). Just showing up: Supporting early literacy through teachers' professional communities. *Phi Delta Kappan*, *82*(9), 675–680.

Mossensen, J., Hill, P., & Masters, G. (1993). *Test of reading comprehension*. Melbourne: Australian Council for Educational Research.

Murimba, S., Nzomo, J., Keithele, M., Leste, A., Ross, K., & Saito, M., et al. (2002). *Monitoring the quality of education for all: Some examples of different approaches used by The Southern Africa Consortium for monitoring educational quality*. 20. Paris, France: IIEP, UNESCO (International Institute for Educational Planning).

Perkins, D. N. 2005. *Understanding, thinking, and education*. Workshop held at Bialek College, Melbourne, April.

Phillips, G., McNaughton, S., & MacDonald, S. (2004). Managing the mismatch: Enhancing early literacy progress for children with diverse language and cultural identities in mainstream urban schools in New Zealand. *Journal of Educational Psychology*, *96*(2), 309–323.

Pool, B., Parkin, C., & Parkin, C. (2002). *PROBE* (2nd ed.). New Zealand: Triune Initiatives.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Paedaogiske Institut.

Richardson, V. (1990). Significant and worthwhile change in teaching practice. *Educational Researcher*, *19*(7), 10–18.

Robinson, V., & Lai, M. K. (2006). *Practitioner research for educators: A guide to improving classrooms and schools*. Thousand Oaks, CA: Corwin Press.

Rowe, K. J., & Hill, P. W. (1996). Assessing, recording and reporting students' educational progress: The case for 'Subject Profiles. *Assessment in Education*, *3*(3), 309–352.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–406.

Tayler, C. (2007). *Challenges for early learning and schooling*. Education, Science & the Future of Australia: A public seminar series on policy. University of Melbourne, Woodward Centre, 23 July.

Thurstone, L. L. (1959). *The measurement of values*. Chicago: University of Chicago Press.

Timperley, H. S., & Robinson, V. J. M. (2001). Achieving school improvement through challenging and changing teachers' schema. *Journal of Educational Change*, *2*, 281–300.

VCAA (Victoria Curriculum and Assessment Authority). (2003–2005). *The Achievement Improvement Monitor*. Melbourne: Author.

Vygotsky, L. S. (1986). *Thought and language*. Boston: MIT Press.

Wright, B., & Masters, G. (1983). *Rating scale analysis*. Chicago: MESA Press.

Wright, B. D., & Stone, M. (1979). *Best test design*. Chicago: MESA Press.