The comfort of competence and the uncertainty of assessment

Patrick Griffin

Paper presented at the Hong Kong School Principal's Conference

Hong Kong Institute of Education,

HK, March 19, 2004.

Competence assessments encourage and entice educators to draw 'can do' conclusions about pupil learning. It is common to describe pupils' progress in terms of things they are now able to do that they could not once do, and we commonly use 'can do' statements to describe competencies: 'can add, subtract, read…. '. There is seduction to such statements and we are drawn into believing that these competences can be assessed by simple observation of people performing specific tasks.

Such an approach is based on a belief that certainty can be attained in describing human ability and development. In this paper I am arguing that a *probabilistic* model of competence can link three fundamental approaches to teaching and learning and provide an appropriate framework for reporting.

This is especially important when goals of education include such things as creativity, teamwork, communication and other somewhat difficult attributes to observe in action. What is clear is that there are specific prerequisite matters that need to be attended to in developing a competency approach to assessment.

1. Specify with greater clarity the desirable *outcomes* of education programs. These outcomes must include a broad range of cognitive, interpersonal, communication and higher-order skills and will need to address the kinds of general competencies identified in national goals of schooling such as those recently espoused in Hong Kong and Australia.
2. Specify *frameworks,* which make clear what is meant by developing competence or achievement. These need to be used as frames of reference for interpreting assessment and identifying how to improve learning.
3. Systematic assessments are needed to supplement the need for reporting in a meaningful way to parents and other stakeholders in education.

Inference is fundamental to all assessment. We are required to 'leap' to the unobservable (an individual's underlying trait) from the observable (performance in particular situations). Rather than attempting 'can/cannot do' conclusions about outcomes, the purpose of assessment is more likely to be to infer a pupil's level of achievement on a developmental continuum that itself represents a manifestation of an underlying trait. The purpose

generally will not be to make ‘can do’ statements in relation to learning. The developmental model of competence takes a probabilistic approach to describing learning development.

Measurement theories and specialists

In the 1960s two developments provided new approaches to the interpretation of assessments. Later in the 1980s, when powerful, accessible computing technology became involved, the real value of the new approaches became apparent. A Danish mathematician Rasch (1980) developed the concept of underlying growth continua, or latent traits. He reasoned that the nature of these traits could be defined by the tasks that pupils performed; if the tasks were to be arranged in order of their increasing amounts of attribute required (capacity), then the nature of the trait was defined by the nature and order of the tasks. Development of learning or competency could be traced by progress along the trait or growth continuum.

In another development, Glaser (1963) put forward the notion of criterion-referenced testing. This, like Rasch’s approach, also described performance and development in terms of the nature and order of tasks performed. Initially, when criterion-referenced interpretation of assessment was used, the observation was referred (or compared) directly to a single, fixed level of achievement or prespecified criterion. If this level of performance was demonstrated, it was interpreted in terms of either mastery or nonmastery, referenced to a single cut-off score. Only one threshold was used. Glaser (1963) originally used the term ‘criterion’ to refer to a defined domain (area) of content or behaviour to which the test items are referenced. Despite its attraction in separating the individual’s performance from those of his or her peers, the can/cannot or mastery/non-mastery interpretation can and did, for more than a decade, reduce the assessment to a level of trivia or to a checklist of unrelated and non cohesive sets of skills that were of little use in instructional terms. It led Glaser who had defined criterion referencing in 1963, to expand on his original statement and clarify the purpose for criterion referencing. Glaser advised that the ‘mastery/non mastery’ tasks should be ordered in coherent sets that lead to an overall interpretation of proficiency or competence.

Glaser (1981) argued that criterion referencing should *...encourage the development of procedures whereby assessments of proficiency could be referred to stages along progressions of increasing competence. (p.935)* Competency development is therefore described and interpreted in terms of a progression of tasks, or stages of developmental competence. There was also no need to define tasks as having only one outcome, approach or solution, and there was no need to restrict the tasks to paper-and pencil exercises that are scored in a predetermined way. Judgment can be used in interpreting performance on more complex tasks, which can be ordered in terms of the competence required to undertake them. The performance then can be interpreted in terms, not only of the tasks undertaken, but also of the manner in which they were completed. When defined this way, criterion referenced interpretation involves the description of the individual or group performance in terms of the tasks undertaken and the behaviors displayed. A criterion then becomes a threshold on the developmental continuum rather then the domain of content.

The person's performance and the task are both interpreted by their relative position on a continuum. This has important implications for teachers and for curriculum developers. It also helps others to share the teachers' insights into pupil learning and the context in which the learning tasks have to be completed. Increasingly it is common to have easily understood and available progressions of increasing competence developing into a communication framework for all parties in the criterion referenced assessment.

Two other developments have made criterion referenced interpretation procedures easier to use. The first is a generalization of the idea of testing. Instead of being seen as paper-and-pencil exercises that predominantly focus on cognitive tasks, tests can now be represented as tasks or procedures performed under specified conditions (Black 1987; Griffin 1987). The second is the availability of powerful microcomputers, which allow analyses that could not be carried out some forty years ago when Glaser first formulated criterion referencing.

Measurement in education and psychology is not as readily accepted as is measurement in physical terms. Physical measurement is generally considered to be independent of the measuring instrument. For example, the length of an object does not depend on which kind of ruler is used. The change from measuring in inches to measuring in centimetres does not change the length of a line. If different rulers are used, each with a different unit of measurement, it is a simple transformation to convert the measure from one ruler to that

from another without changing the length of the object. This characteristic is called the specific objectivity of measurement. Attempts to develop systems of cognitive and educational measurement that emulate the features of physical measurement are not new. Thurstone (1925) and Thorndike (1929) proposed a measurement system for the assessment of mental ability and development based on a scaling procedure applied to discrete-point test items. Various others worked on developing mathematical models of measurement and Rasch (1960) developed the simplest of a family of latent-trait models that use the statistical characteristics of a group's response to a set of items to determine the measurement scale.

These methods have been collectively grouped under the heading 'latent trait models'. The traits are hidden (or latent). By their construction and interpretation, they help us to interpret or understand observations; there is no implication that they exist in any physical or physiological sense or that they cause the behavior they describe. This is also true of competency definitions: the descriptions of competency help us to observe and interpret behaviors exhibited by people in performing tasks. The competency is defined and constructed by us to help us discriminate between pupils on the basis of skill. Each competency is a description that is used as an aid to observation and interpretation; it does not exist in its own right. When we apply statistical theory to data on these observations, we take the construction one step further. Now we have a statistical representation of a verbal description to assist in interpreting observations of learning. *Latent traits are statistical constructs derived from empirical relationships among records of observations of peoples' behaviours. So too are empirically defined competences.*. Analyses of performance information focuses on the nature of the tasks performed, the difficulty of the tasks and the way they separate those with more of the trait from those with less of it (task discrimination).

Unlike traditional tests, the tasks do not necessarily have single, correct outcomes and, as they become more complex, criteria defining thresholds or levels of performance can be used. The theory is concerned with the relationship between the demands of a task and the capacity of the person to perform it. There is no restriction on the nature of the task and in the most general of the Rasch models (Linnacre 1990), there are very few restrictions on the scoring procedures. The task can be a test question, a set of multiple choice items, an essay, a performance, a speech, a product produced in class, an artistic rendition, a folio, a

driving test, the dismantling and reassembling of a motor car engine, the building of a brick wall, a haircut given to a client, or whatever is related to some attribute of interest. The attribute could be ability, an attitude, a physical performance, a procedure, an interest, a set of values or a generalised competence in an area of learning. An advantage of the approach is its links to the arguments separately posited by Vygotsky, whose research may be interpreted as being driven by questions about the development of human beings and about the role that formal education plays in the process. The concept of *internalisation* provides the basic strategy for teaching higher order thinking and competence in inquiry.

It is in the context of this broader conceptualisation of learning that we may interpret Vygotsky's construct of the Zone of Proximal Development (ZPD) - the zone in which an individual is able to achieve more with assistance than he or she can manage alone. Two expositions of the ZPD in his published work arose from two different immediate concerns: the assessment and placement of children who were "learning disabled" and the role of instruction in the development of scientific concepts (Vygotsky, 1934, 1962/1986).

Simply put, the ZPD is

> *"a state of readiness in which a pupil will be able to make certain kinds of conceptual connections, but not others; anything too simple for the pupil will quickly become boring; anything too difficult will quickly become demoralising"*
>
> (Burbules, 1993, p122)

From the point of view of the teacher-pupil relationship, the teacher needs to be able to identify "the state of readiness" of the pupil in the context of the domain of learning being mastered. These ideas have affected pedagogical practice particularly as evidenced in studies on 'modelling' and 'scaffolding'.

Interestingly, the measurement theories of Rasch are also consistent with Vygotsky's approach. Glaser's words "*stages along progressions of increasing competence"* are of immense importance in competency assessment, whether it is in assessing Basic

Competencies or any other form of competency. Criterion referenced interpretation now incorporates Vygotsky's ZPD and this is formalised when the two ideas are linked to item response theory applications in assessment. Combining these ideas directly links the position of a person or an item on a developmental continuum (as shown in a variable map) to an interpretation of what a pupil, or groups of pupils, can learn (with assistance), rather than focussing on a score or the performance relative to a percentage or a group. It gives a substantive interpretation to the measurement for reporting purposes rather than relying on a score or grade. The procedure gives meaning to test scores and enables them to be interpreted in terms of Glaser's levels of increasing competence.

A weakness of the method is its reliance on a small number of specialists in item response or latent trait analysis. The underlying latent traits are first examined using a system called variable maps as shown in Figure 1.

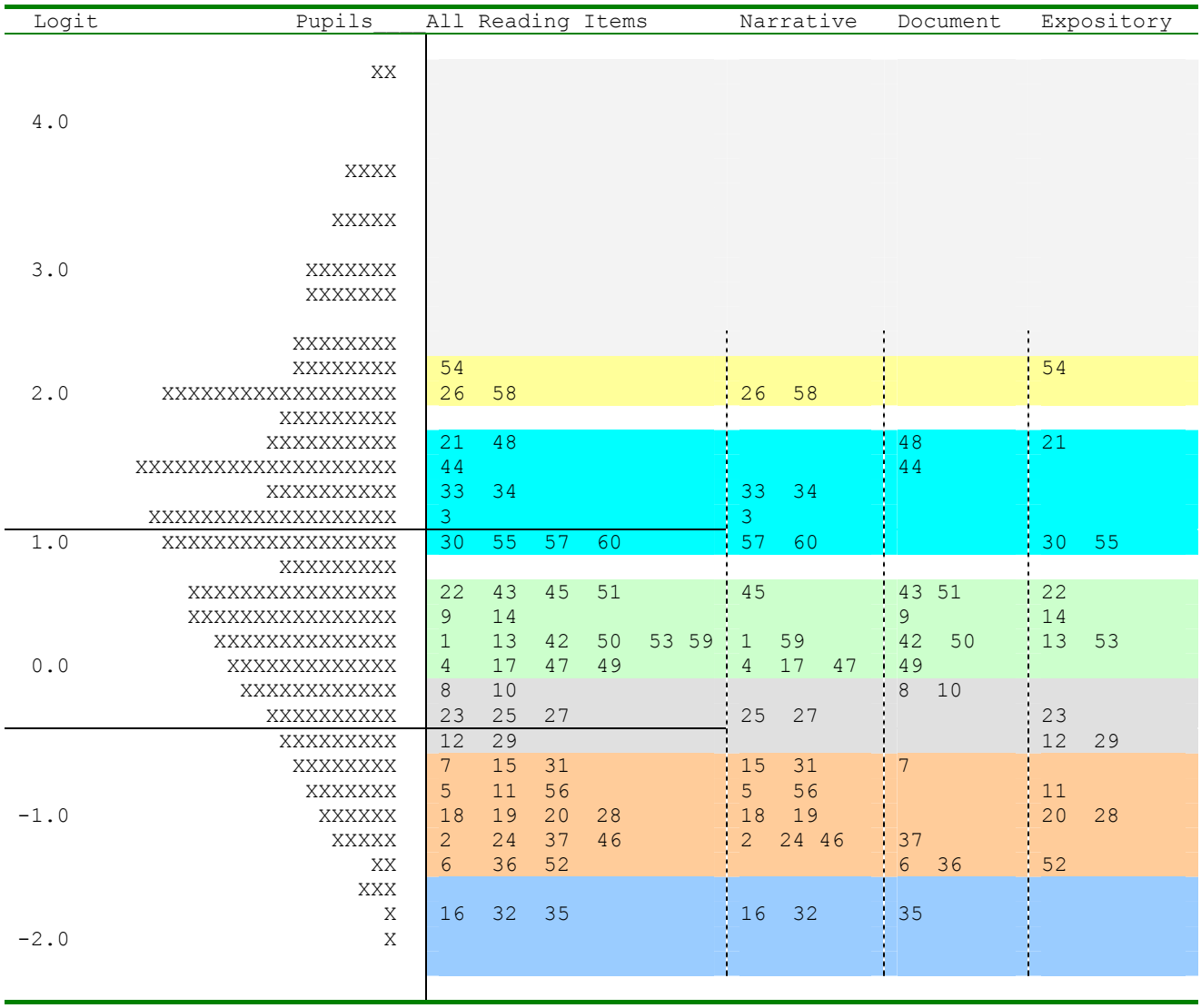

Figure 1: Variable map of a 60 item reading test

These are generated using any number of Rasch model computer software packages. It can be seen that several test items group together at different points along the scale and a major question is whether these clusters can be interpreted as having something in common. Each item is reviewed for the skills involved in responding to the item and is a matter of substantive interpretation. The process requires an understanding or empathy with 'how the pupils think' when they are responding to the items. Experienced teachers are very good at this and those dealing with language instruction, and who are accustomed to dealing with a marking scheme, can readily interpret the levels on a developmental continuum underpinning a test.

A variable map shows that items can be grouped according to similar difficulty levels. Given that the ability of the pupils is matched to the difficulty of the items and the items and pupils are mapped onto the same scale, the pupils can be grouped within the same 'ability' range adjacent to the items that have similar difficulty levels. This grouping of items (and pupils) identifies a kind of 'transition point', where an increase of item difficulty is associated with a change in the kind of cognitive skill required to achieve a correct answer.

When ability and difficulty are equal the odds of success are 50/50. From this it can be deduced that, if the pupil were to improve a little, he or she would have a better than even (50/50) chance of succeeding on items in this group. 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. This leads to an understanding of the kinds of skills being demonstrated by pupils at each level on the continuum. Moreover, the odds of 50/50 at the transition points could be linked to a change in the required cognitive skill and this could be directly translated into an implication for teaching. It is in fact the zone of proximal development defined by Vygotsky. Discussions with curriculum specialists need to take place to identify the kind of instruction needed to progress the pupil on the variable. A summary description of these skills can then be assigned to each item and pupil group.

The first point (item grouping) is 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 is sometimes described as a Rasch-like manner because it is also a requirement of the Rasch model analysis. The second point (labelling the skills) is based on conceptual rather than on statistical grounds and is based on a skills audit.

To achieve this outcome the difficulty measures of the items are sorted in increasing magnitude. Each assessment task also needs to be analysed for the underpinning cognitive skill involved in obtaining the correct answer. The results of these analyses for a reading test are presented in Table 1.

Table 1: Skills audit for each of the 60 Pupil reading test items

Item #	logit	Cognitive skill underpinning the correct response
41	4.2	Link a concept to a visual stimulus and bring outside knowledge to the solution
54	4	Combining several ideas and using outside knowledge (format-all of the above)
58	3.8	Combining several ideas, requiring interpretation beyond text level
26	3.8	Combining several ideas (format-all of the above)
21	3.5	Combining several ideas and using outside knowledge
48	3.4	Understanding figurative meaning of word (format-negative question)
44	3.3	Understanding author's main purpose on the basis of the title
34	3.1	Requiring interpretation beyond text level, unfamiliar topic
33	3.1	Deducing meaning from context
3	3	Combining several ideas (format-all of the above)
55	2.8	Understanding main idea, choosing a title
30	2.7	Inferring meaning from context (format-negative question)
60	2.7	Understanding figurative meaning
57	2.7	Locating specific information from text
40	2.7	Locating specific information from text
43	2.4	Locating specific information from text
38	2.4	Locating specific information from text (too many details in long options)
22	2.4	Locating specific information from text
45	2.3	Locating specific information from text
51	2.3	Locating specific information from text
9	2.3	Integrating reading and math skills
39	2.2	Locating information from text & illustration
14	2.2	Locating information from text (understanding signal words-"prediction" in the stem)
50	2.1	Inferring meaning from context (option d attracts some above average pupils)
53	2.1	Locating information from context
59	2.1	Understanding author's main purpose
1	2	Match exact words and paraphrase from Chinese origin
13	2	Locating information from text (format - negative questions)
42	2	Locating information from text
47	1.9	Understanding implications
4	1.8	Understanding implications
49	1.8	Locating information from text
17	1.8	Understanding meaning of sentences
10	1.7	Understanding meaning of vocabulary
8	1.6	Integrating reading and math skills
27	1.5	Understanding meaning of sentences
25	1.4	Locating information from text
23	1.4	Locating information from text
12	1.4	Locating information from text (format - negative questions, using background
29	1.3	Locating information from text
31	1.2	Locating information from text
15	1.1	Understanding meaning of words
7	1	Locating information from text
56	1	Locating information from text
11	1	Locating information from text
5	0.9	Understanding author's main purpose
18	0.8	Locating information from text
19	0.8	Understanding relationship between events in text
28	0.7	Locating information from text
20	0.7	Locating information from text
37	0.7	Locating information from text
2	0.6	Locating information from text
46	0.6	Understanding relationship between events in text
24	0.5	Understanding meaning of word
52	0.4	Locating information from text
6	0.4	Locating information from text
36	0.4	Matching word and visual stimulus
16	0.1	Exact match of text with adjacent text
32	0	Match exact words and paraphrase
35	0	Matching word and visual stimulus

In this example, the difficulties of the test items (logits) were also plotted in increasing order of difficulty and the sets of items were examined to identify specific clusters or groupings. The two criteria described above were used. First, there have to be identifiable sets of items and these sets need to have a common substantive interpretation of the underpinning skill. Grouping items on the variable map is a first step, but it is imprecise because of the constraints of printers and line feeds and this may place some items with different difficulty on the same physical line merely because of a hardware restriction. Nevertheless it is a good first step, as an inspection of the variable map can often identify broad categories and clusters. The chart in Figure 2 illustrates where the difficulty of items changed. The question then arose that if the difficulty increased for sets of items, did the nature of the underpinning skill also alter? The two sets of information were explored in unison. Natural breaks in difficulty were identified and then the items and the cognitive descriptions were examined to determine if a set with a common substantive interpretation could be found. A panel of curriculum specialists joined the item writers for this exercise. Together they identified the breaks in the variable and offered the substantive interpretation of the levels of competence. These have been presented in Figure 3.

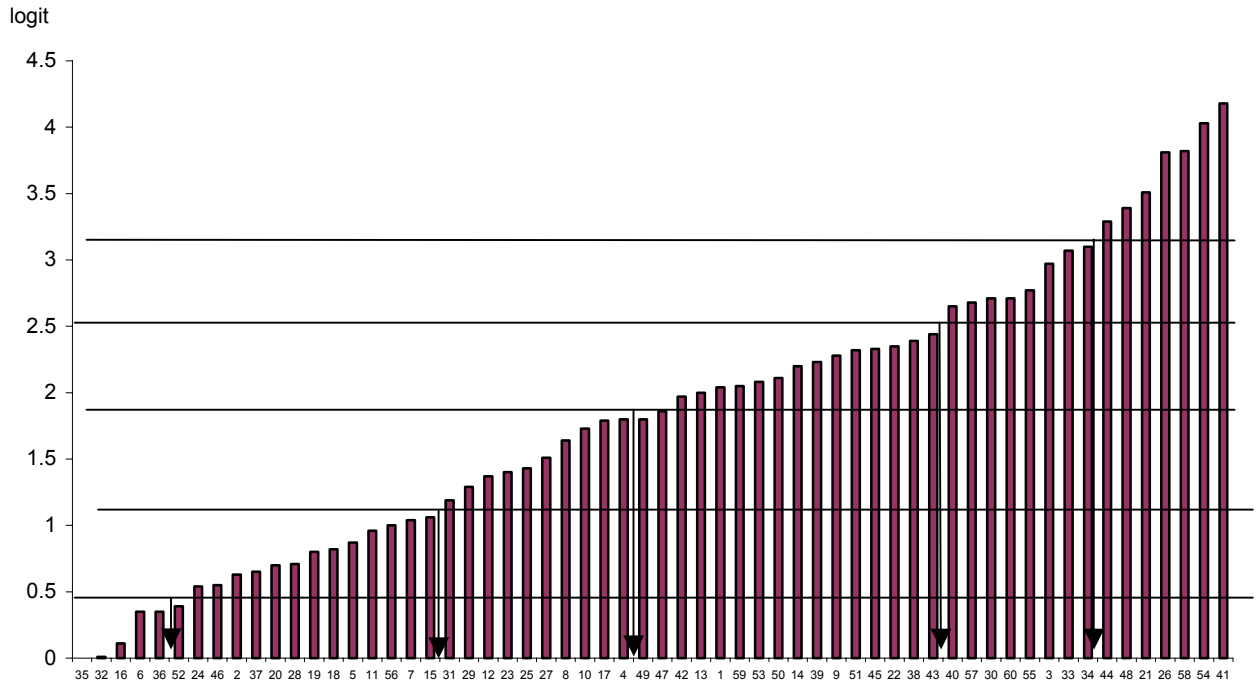

Figure 2 Relative difficulties of reading items and cut points for competence levels

The clusters of items was then interpreted and the following broad levels of competence were identified.

Reading Skill Levels	
Level 1	Matches text at word or sentence level aided by pictures. Restricted to a limited range of vocabulary linked to pictures
Level 2	Locates text expressed in short repetitive sentences and can deal with text unaided by pictures. Type of text is limited to short sentences and phrases with repetitive patterns.
Level 3	Reads and understands longer passages. Can search backwards or forwards through text for information. Understands paraphrasing. Expanding vocabulary enables understanding of sentences with some complex structure.
Level 4	Links information from different parts of the text. Selects and connects text to derive and infer different possible meanings.
Level 5	Links inferences and identifies an author's intention from information stated in different ways, in different text types and in documents where the message is not explicit.
Level 6	Combines text with outside knowledge to infer various meanings, including hidden meanings. Identifies an author's purposes, attitudes, values, beliefs, motives, unstated assumptions and arguments.

Figure 3: Interpretation of the reading levels from the analysis of reading test item sets

However not everyone is able to use the information at this level. Still more work is needed…

"These test results, competency levels and curriculum materials are interesting, but how are they used?"

That teacher's remark illustrated the importance of professional development linked to an assessment and reporting program. Both pre-service and in-service teacher education programs will be central and new teaching methods will be needed to accommodate the links between criterion referencing, item response modelling and the zone of proximal development. Before this can be successful, teacher training programs will need assistance to broaden their program beyond the focus on the central curriculum model and the related teaching plans and to incorporate methods of developing the classroom teachers skills in combining these three theories. Teaching and assessment strategies need to be merged and specific teaching strategies developed for each zone of proximal development.

Teachers ALSO need help to make the transition. Team teaching, cross level tutoring, teacher mentoring for less confident and competent teachers are essential strategies. If, for example, Hill and Crevola's strategy (1999) were implemented such that the weakest

teacher in a school can be mentored and helped to improve to even the average level for the school and if this translates to pupil achievement, the impact on the system would be enormous.

For pupils, placement and diagnostic testing need to be much more widely used provided that their interpretation leads to targeted intervention for all pupils. The image of diagnostic testing is that it is for remedial teaching, but this needs to be broadened to targeted instruction for all levels of reading and mathematics. This leads automatically to the next section in this discussion- the link of teaching to learning outcomes and competency levels.

Instructional implications

Overhauling the instructional agenda to cater for different levels of achievement and learning style is a large task and one that will take time. It is more than the procedural tasks outlined in the above discussion. Each of these is constructive but none of them answer the teacher's question.

Early readers need activities like a listening centre, pocket charts, enlarged texts (such as those being developed through the UNICEF in its global approach to Education for All (EFA)), word walls, poem boxes, buddy reading. Higher level readers are encouraged by literature or book circles, directed reading, semantic webs, sketches and other innovative approaches.

Where classes have pupils over a wide range of language competence the whole class approach is inappropriate. The example provided here focuses on reading instruction and the examples are taken from the Literacy Profiles by Griffin, Smith and Ridge (2001) or by Griffin, Smith and Martin (2003). A similar analysis of teaching strategies could well be done for the range of competency levels in mathematics, and this could well be a part of the training modules developed elsewhere.

Derived competency statement	Possible teaching activities
Level 1: Matches text at word or sentence level aided by pictures. Restricted to a limited range of vocabulary linked to pictures	Reading conferences, logs kept by pupils, shared reading, retelling, drama activities based on reading. Books sorted into difficulty levels and pupils practice by reading aloud, reading to other pupils, parents, use 'take-home' books, reading by the teacher to pupils, reading simple repetitive language pattern picture-rich reading texts repeatedly alone and with other pupils,
Level 2: Locates text expressed in short repetitive sentences and can deal with text unaided by pictures. Type of text is limited to short sentences and phrases with repetitive patterns.	Reading to, with and by the pupils in the class reading centre. Shared reading with other pupils, parents, use 'take-home' books, reading by the teacher to pupils, guided reading and predicting stories with simple repetitive language pattern and picture-rich reading texts, repeatedly alone and with other pupils, Reading logs, sustained silent reading, retelling, running records readers' theatres and creative drama.
Level 3: Reads and understands longer passages. Can search backwards or forwards through text to for information. Understands paraphrasing. Expanding vocabulary enables understanding of sentences with some complex structure.	Comparing books and stories, identifying features, exploring common patterns using reading circles, sustained reading activities, discussions with other pupils and parents and recording reading logs and discussions; role plays, portfolios, individual reading conferences, guided reading programs at the individual pupil level; retelling; and links to writing instruction, collecting stories and other reading materials from the community. Small group activities and reading centres;
Level 4: Links information from different parts of the text. Selects and connects text to derive and infer different possible meanings.	Guided reading, small-group reading activities reading circles, reading logs reading materials from community and from non fiction, shared reading focussing on strategies for expository texts collected from a range of sources, and related to a range of curriculum learning areas; reading aloud, following directions, keeping reading diaries, writing letters; individual reading conferences, reading logs and response journals
Level 5: Links inferences and identifies an author's intention from information stated in different ways, in different text types and in documents where the message is not explicit.	Reading targets in terms of the number and range of texts and text types. Non fiction should be extensively used, unit and topic research activities as individual and group activities with work set for research at home and in the community. Critical analysis of text materials and evaluation of writers style and effect of style.
Level 6: Combines text with outside knowledge to infer various meanings, including hidden meanings. Identifies an author's purposes, attitudes, values, beliefs, motives, unstated assumptions and arguments.	Shared, guided and independent reading of a broad range of text types and from a range of sources. Retelling and reading circles predominate in teaching styles and activities, sharing insights, clarifying intentions analysing and evaluating texts using a range of criteria such as style, clarity, impact on the reader and so on. Book clubs, think and know charts based on reading, drawing conclusions from a range of texts, following directions, drama workshops, reading and writing discussion groups, literature response portfolios text cohesion analysis sessions.

Figure 4: Developmental levels (ZPDs) and intervention strategies

At a system level, this has further implications. If changes in teaching are required then it is also true that changes in resource provision are involved and this is turn has implications for policy development.

Teachers need assistance to use the theories in their teaching without the specialised theoretical and computer training that all this implies. A simple procedure that we have developed for teachers (Griffin, 2001) encompasses all of these theories and consists of the following steps. Teachers take just a few minutes to grasp the approach using the following steps to adapt it to their teaching.

1. Describe the teaching and learning sequence that would facilitate learning through these three levels and devise a series of tasks that would track through the three levels.
2. Define the levels that are expected to develop.
3. Develop a set of specifications for the assessment task(s)
4. Identify the steps of the components of each task.
5. Anticipate, for each step, different quality of performance from different pupils.
6. Define codes for the different quality performances.
7. Put the codes on a continuum of increasing level of performance.
8. Identify clusters of codes and define overall levels of performance.
9. Interpret the clusters of codes using the performance rubrics
10. Define the characteristic knowledge and skills for each level and interpret these in terms of readiness to learn.
11. Identify intervention strategies for each level
12. Set target levels for mastery and/or competence
13. Design a report that indicates achievement, readiness to learn and intervention
14. Adapt the report to illustrate performance on a criterion, standards and norm reference scales, as well as competency scale, and a norm referenced interpretation.

There is no need for sophisticated computing or large scale survey testing work. The work of Rasch, Glaser and Vygotsky can all be incorporated into the teaching and assessment cycle of a teacher using some simple planning tools listed above.

Reporting

The final stage is about telling parents, pupils and other teachers about the progress along the continuum. What does it all mean? A form of reporting that enables all three theories to be demonstrated is illustrated in the following figures.

The report for the pupil is then generated to show how the performance is interpreted using the rubrics for segments as defined in the tables above. The figures below show both the individual pupil and the class reports.

We assume that a criterion referenced analysis and interpretation needs a criterion referenced report format for the pupils and parents. There are two popular ways in which

this is done. The first is an item level content analysis and was pioneered by the ACER. The second was pioneered by the Assessment Research Centre at the University of Melbourne and is called a profile approach. Both are Australian models of test interpretation. The first is an item analysis and a report on each item and success rates. The second is a profile approach whereby stages of increasing competence are defined and used for reporting procedures. The Melbourne University Centre now combines these two approaches and the combined method is to be published for the first time in 2004 in a World Bank report (Griffin, 2004). The advantage of the item level analysis alone is that the pupil receives a detailed report on the performance in the examination. The disadvantage is that, unless the subsequent examination is similar or identical, it can promote learning not conducive to a developmental progression. It can also lead to a trivialising of learning.

The advantage of the second approach is that it transcends the actual content of any one test and allows the pupil to study the general construct being measured. It also indicates what standards are being used.

The criterion report is shown in Figure 5. When this is linked to an item content analysis the pupil can be given information about the variable being measured, the criterion levels and the item analysis.

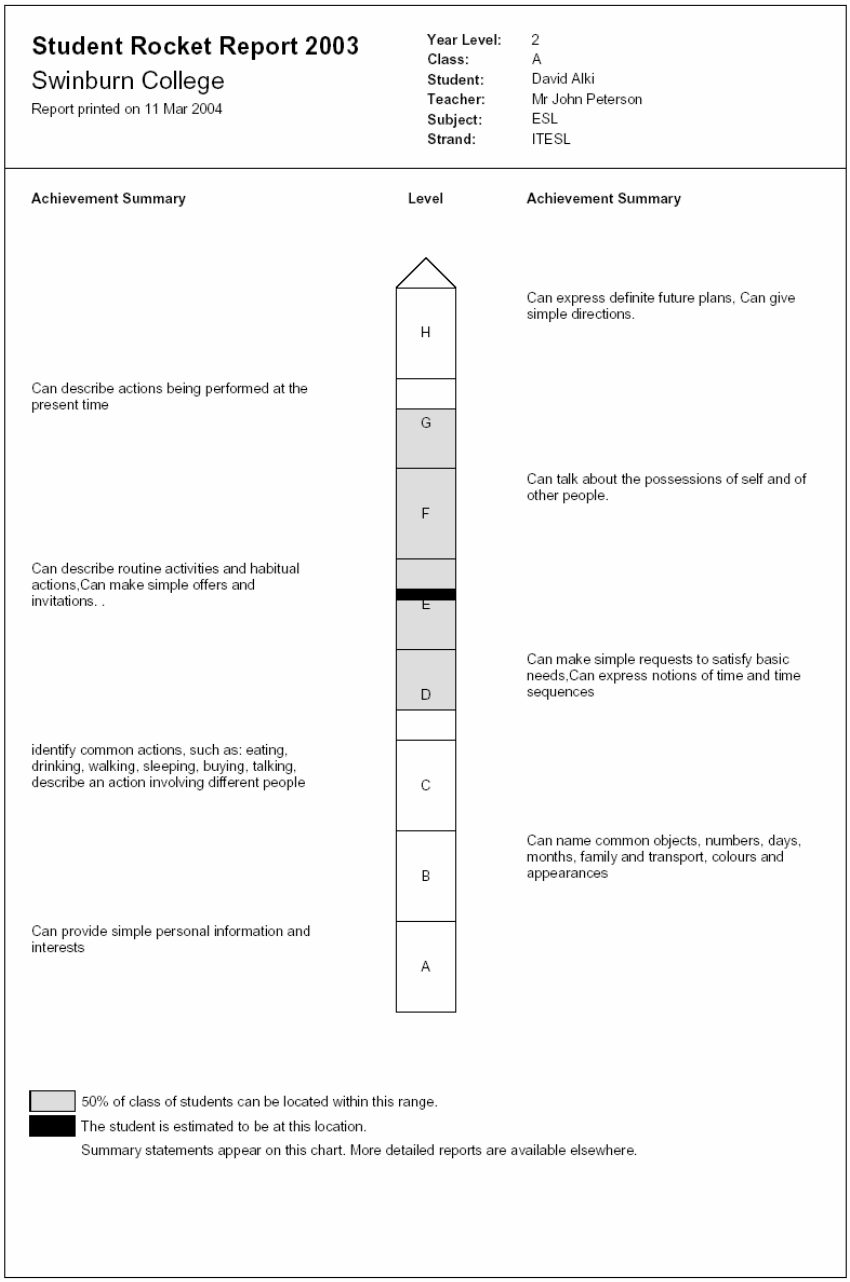

Figure 5: an example of a criterion referenced report format at an individual level.

Figure 5. illustrates a way of reporting at an individual pupil level on a criterion scale. The central axis of the diagram presents levels of increasing ability. This could be labeled with just letters or it could be labeled with the Rasch logit scale. To the right and left of the central spine of the diagram are descriptions of levels of increasing competence depending on an analysis of items that might be located at around those regions. The grey region in the middle of the central column indicates the inter- quartile range of pupils (i.e. the bottom of the grey box would be the 25th percentile the top of the grey box would be the 75th percentile). The dark line in the middle indicates the pupil's level of achievement

on this test. Interpretation of this chart shows that by reading to the left and right of that line, it is possible to identify the skills that have been developed below that line and the skills not yet able to be demonstrated above that line. It also illustrates where the ZPD for the specific pupil is located.

In an expansion of this diagram in Figure 6 a class group is examined. This would be reported to the teacher not to pupils or parents, but it presents two pieces of information. Across the top of the chart is a listing of the skills and knowledge that have been demonstrated through responses to items in the examination.

The left vertical axis presents pupils' names. The grey shaded region in the centre of the chart is the inter-quartile range as presented before in the earlier chart. The histogram or the bars within the graph are presented as a dark grey and black columns. These represent measures at time 1 and time 2. Plus, as pupils may undergo pre and post assessment in an action research agenda, it is possible to present both the pre and the post performance of the examinees in terms of the criterion referenced scale and developmental (ZPD) level demonstrated on the test.

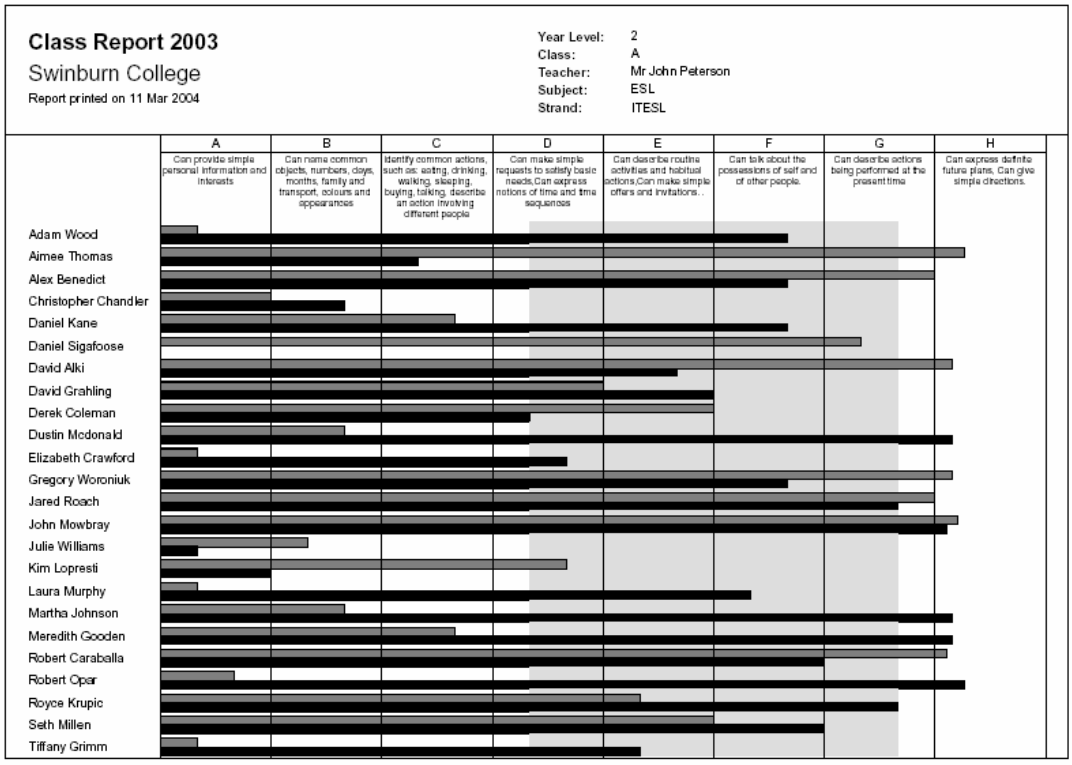

Figure 6: Criterion referenced group report.

Figure 6 shows how pupils can be monitored over time and how many attempts at the examination can be monitored over time and their relative ability level can be monitored. The opportunity to give feedback is important.

Given the issues raised in this paper, several conclusions can be drawn.

1. Definitive statements about pupil performance and capabilities cannot be made. It is only a probabilistic context.
2. Strategies need to be developed to enable probabilistic statements to be made about learning
3. Item response modelling, criterion referenced interpretation and zones of proximal development come together in a single approach to teaching learning and assessment.
4. Assessment is best used to identify the intervention strategy associated with the ZPD.
5. Identification of the intervention strategy and the ZPD implies that different teaching strategies are needed for each ZPD and this in turn implies differential allocation of resources.
6. When this is applied at a system level the allocation of resources has implication for curriculum and resource policies.
7. Teachers can readily adapt the logic of the probabilistic models of assessment to develop and implement the criterion referenced, IRT and ZPD approaches to teaching assessment and learning in their classroom without large scale adoption of sophisticated computer models.
8. A criterion referenced and ZPD approach to teaching and assessment demands a reporting morel that enables this to be communicated to a range of stakeholders and reports that illustrate scores, grades or comments unrelated to the criterion, latent trait and proximal approach are of limited use.

References

Burbules, N. (1993) *Dialogue in Teaching*. New York: Teachers College Press.

Glaser, R. (1963). "Instructional Technology and the measurement of learning outcomes: some questions." American Psychologist **18**: 519-521.

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

Griffin, P. (2004) Chapter 2. The construction and calibration of tests. in Volume 3. Postlethwaite, N., Griffin P. et al… Achievement Levels of Vietnamese primary students. Washington: World Bank.) (In press)

Griffin, P. Students, Take your Marks! Get Set - Learn! (2001) Keynote address at the New South Wales Chapter Conference of the Australian Council for Educational Administration, Mudgee, 2001).

Hill, P., and Crevola, C.: *"The Role of Standards in Educational Reform in the 21st."*ASCD Yearbook, 1999.

(Rasch 1980)

Griffin, P. Smith, P. and Martin, L. (2003). Profiles in English as a Second Language. Portsmouth, N,H.: Heinneman.

Griffin, P. (2001). ALPS: Software Support for the Literacy Profiles. Portsmouth, Heinemann.

Griffin, P., P. Smith, et al. (2001). The Literacy Profiles in Practice: An Assessment Approach. Portsmouth, Heinemann.

Linnacre (1990)

Postlethwaite, N., P. Griffin, et al. (2002). Reading and Mathematics Achievement of Vietnamese Children (in press). N. Postlethwaite. Hanoi, The World Bank. **2:** 103-147.

Rasch, G. (1980). Some probabilistic models for the measurement of attainment and intelligence. Chicago, MESA Press.

Thorndike, L. 1929 'A method of scaling psychological and educational tests. Journal of Educational Psychology, vol. 16, pp. 433-49

Thurstone, E.L. 1925 *The measurement of intelligence* Columbia University, New York

Vygotsky, Lev (1986). Thought and Language. Cambridge, Massachusetts: The MIT Press

Vygotsky, Lev (1934). Thinking and Speaking. Cambridge, Massachusetts: The MIT Press. Written 1934: Edited and translated in 1962 by Eugenia Hanfmann and Gertrude Vakar