Can we trust our teachers, their tools and techniques?

Sivakumar Alagumalai.
School of Education. University of Adelaide. Australia

Abstract:

There have been positive directions in formative assessment and school-based assessment. These assessments provide students opportunities to highlight their learning in a developmental and progressive manner in authentic settings. Learners are empowered to indicate, through their work, both the heuristics of the learning and content mastery.

Teachers (or raters!) in schools have the arduous task of examining a student's work (against a list of capabilities statements or through a set of criteria articulated in a rubric) and making objective judgements. Scores are assigned to various activities and tasks, which are then aggregated for a grade and reported to the students and relevant stakeholders.

This paper highlights the problems with raw scores, its aggregation and the effect of raters (or teachers) on the scoring and grade assignment processes. It also discusses the challenges associated with rubrics and judgements, grade inflation and comparability. The use of Rasch and multilevel techniques is highlighted. The principles underlying objective measurement are included.

Keywords:

teachers, assessment, formative assessment, rubrics, distractor differential functioning, rater bias, Rasch model, cross-classified multilevel measurement

---

Introduction

Teachers play a pertinent role in society and provide through schools the transition-base between family and the broader community and society. “There is probably no profession as exciting and as personally rewarding as that of teaching. Each day presents anew the opportunity to enrich the lives of others and one’s own in the process” (Dunn, 2005, p.1). Thus, a major goal of teachers, schooling and the curricula is to prepare students to flexibly adapt to new settings and problems, and successfully transfer and apply their learning. Bransford, Brown and Cocking (2000, p.236) argue that “effective teachers attempt to support positive transfer by actively identifying the strengths that students bring to a learning situation and building on them, thereby building bridges between students’ knowledge and the learning objectives set out by the teacher.”

Wink (2005, p.166) highlights the ‘demands and needs’ of students of the twenty-first century and the contributions teachers can make. These demands include the “need for bilingual and biliterate students who love to read, can reflect critically, and live their lives with passion and action; need collaborative, lifelong learners who are responsible for their own learning and understand that it comes from their lived experiences; need students who can generate new knowledge and apply it in unknown ways; need students who can write and rewrite their world from a pluralistic perspective, students who can pose problems and solve problems with technology; need students who know how to access, interpret, and critically use new and emerging information; need to be able to work in multilingual and multicultural society.”

Hassett (2000) outlines the characteristics of effective teachers who facilitate the learning processes. According to her, good teachers are reflective and know how to live with ambiguity. Effective teachers routinely and systematically think about and reflect on their classes, their students, their methods (techniques), and their materials (tools). This parallels Bransford, Brown and Cocking (2000, p.236) point that teachers acknowledge and understand that students already have relevant knowledge either in line with what the planned lesson is or as alternative conceptions. However, the greatest challenges of teaching stems from the lack of not fully understanding where students are positioned in their knowledge and experiences, and the provision of immediate accurate feedback. Experience, and pedagogical content knowledge of teachers are crucial factors that decrease the dependence on externally developed materials and support, and enhance accurate predictions of both student learning and diagnostics. Thus, assessment, whether formative or summative, should empower teachers to “help students change their original conceptions rather than simply using the misconceptions as a basis for further understanding or learning new materials unconnected to current learning” (Bransford, Brown and Cocking, 2000, p.236).

This paper examines the interactions between teachers with the tools they may have created or used to gauge learning, and the techniques they deploy to provide feedback to students. The paper re-examines the challenge advanced by Dunn (2005, p.1) that “teaching, as a profession still has not reached the status of other profession such as medicine, law and engineering. Whether or not it ever will is largely dependent on the decisions that are made both now and in the future by those in the profession.”

Teachers, teaching and decision-making.

There is an urgent need to link curriculum, instruction, assessment, and standards in a more generative and even transparent way (Bond, 2004). Teachers are encouraged to anticipate the difficulties students will have with various concepts and how to structure and sequence instruction to minimize these difficulties. 'Expert teachers' know the structure of the knowledge in their discipline areas. This knowledge provides them with cognitive roadmaps to guide the assignments they give students, and the assessment they use to gauge student progress. In this way, both students' prior knowledge and teachers' knowledge if subject content become critical components of learners' growth. (Bransford, Brown and Cocking, 2000, p.241).

To optimise the teaching profession, a number of teachers' standards have been drawn up, and generally include the following (Harman, 2001):

- Teachers are committed to students and their learning,
- Teachers know the subjects they teach and how to teach those subject to students,
- Teachers are responsible for managing and monitoring student learning,
- Teachers think systematically about their practice and learn from experience, and
- Teachers are members of learning communities.

It must be acknowledged that teachers are at various levels of proficiency (and competency, if I may include) and that there is no way to predict precisely what the long-term results of classroom teaching will be. Carlsen (1999) acknowledges that teachers have varying pedagogical content knowledge, and decisions about teaching, materials, and assessment will differ. Within-teacher and between-teacher variations and decision-making processes about teaching and assessment thus need careful scrutiny.

There is an urgent need to re-assess what we mean by pedagogy. Wink (2005, p.1) indicates, "Pedagogy is to good interactive teaching and learning in the classroom as critical pedagogy is to good interactive teaching and learning in the classroom and in the real world." Increasingly, educational quality and relevance are defined by reference to students' learning outcomes (UNESCO, 1998, p.48). In moving from a pedagogy that is highly transmission-based, through a generative pedagogy, and to a transformative one, assessment of learning has to be considered seriously so that students are not biased or prejudiced. The pressure on teachers, and their decision-making (or judgement) mounts with increased transparency and accountability. The movement towards monitoring and evaluation of the quality and performance of national education systems has undoubtedly begun to have an impact on the way in which education is regarded both by society at large and by the people directly involved, not least teachers (UNESCO, 1998, p.52).

We can examine the 'products and processes' employed by engineers, doctors and lawyers. Their practices are opened to examination and challenges. As a profession, we are (as always) obliged to make transparent our choice of teaching/assessment tools (materials), and techniques (strategies) for assessment. There is an element of 'unpredictability' in our choices, and should make explicit on what we can control. Alagumalai (2006) argues that any interaction involving behaviour and 'life' adheres to a probabilistic function, at both the microscopic (as in genetic interactions and mutation) and macroscopic (mass migration) levels. Thus, the interactions between teacher and student, teacher and student's work, teacher and teaching tools (and aids) are highly probabilistic. It will be disastrous if parents and stakeholder are kept second-guessing why a student is assigned a particular score and/or

directed through a particular diagnostic pathway. More broadly, curricula developers and policy makers need this assessment information to redesign curricula and general directions in education (Alagumalai, 1996; Pellegrino, Chudowsky, & Glaser, 2001). The next section provides the arguments for examining teachers' tools and techniques, and exemplars of research undertaken.

Interaction of teachers their assessment tools and techniques

Assessment, whether summative or formative, has been acknowledged as the most important contributor to learning (Black, 2004). However, the grading and reporting of student learning have created the greatest controversy among educators (Pollio & Beck, 2000). The design and development of the assessment instrument/task (including Table of Test Specifications), associated assessment criterion (including rubrics), award of raw scores to the various item/task, conversion of raw scores to a grade, and feedback to students have an element of fuzziness and involves teacher judgements. As highlighted by Alagumalai (2006), there is an interaction between the teacher and the assessment 'contents and processes'. The assignment of grades shifts the meaning of assessment into evaluation (Keeves and Masters,1999, pp.14-15). Thus, student evaluation has a value judgement component, and shifts the meaning of traditional assessment away from the learning-diagnostic-remediation processes. The assignment of scores to items and tasks can become problematic and hence contested. Even though there are moderation processes in place, the ambiguities of raw scores and grades need attention. Figures 1 and 2 highlight the challenges associated with the assignment of grades and grade inflation.

A VARIETY OF GRADING SCALES

Here are examples of four professors' grading scales from four departments at Virginia colleges and universities.

	WILLIAM H. ALEXANDER History, NSU	GILBERT YOCHUM Economics, ODU	LARRY SABATO Politics, U.Va.	CRAIG WANSINK Religion, Va. Wesleyan
A	95-100	90-100	96-100	94-100
A-	90-94	-	93-95	91-93
B+	87-89	-	90-92	88-90
B	84-86	80-89	87-89	84-87
B-	80-83	-	84-86	81-83
C+	77-79	-	81-83	78-80
C	74-76	70-79	78-80	74-77
C-	70-73	-	75-77	71-73
D+	67-69	-	72-74	68-70
D	64-66	60-69	69-71	64-67
D-	60-63	-	66-68	61-63
F	59 & below	59 & below	65 & below	60 & below

SOURCE: The professors

Figure 1: Variation in Grading Scale
http://www.hamptonroads.com/pilotonline/special/grades/nw0210scales.html

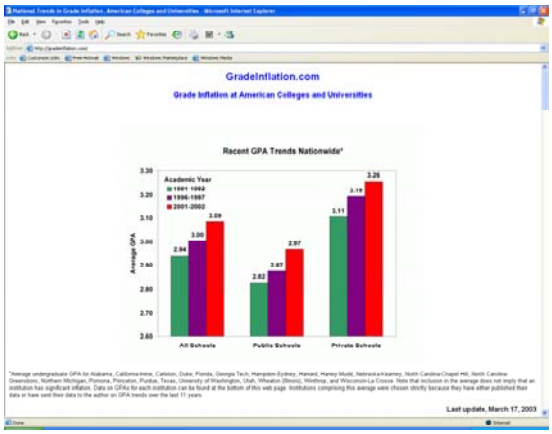


Figure 2: Grade Inflation of GPA
http://gradeinflation.com

This together with the 'rubber ruler' raw scores, irregular intervals and squashed extremes, as highlighted by Wright (1999), flag major concerns of what teachers' judgements on assessment mean. Thomas and Bainbridge (1997) warn that grade inflation is the 'current

fraud'. Johnson (2003) and Kuh & Hu (1999) further indicate the problems with subjective teacher judgement and its implications for grades and the motivation to learn. There is a shift in the way we view assessment of and evaluated students' work, and the demand for examining assessment as a key component in pre-service and in-service teacher education becomes fundamental. The exemplars provided in the next section highlight an urgency to reconsider professional developments of teachers/educators.

Case Study One: 'Faulty' distractors (Alagumalai & Keeves, 1999).

Alagumalai and Keeves (1999) examined the problems that may exist at the item and distractors levels. A number of terms have been used to identify items and distractors that may bias a particular subgroup of test-takers, and include differential performance, differential functioning, and systematic errors to name a few. QUEST, software that utilizes the Rasch model, was used to examine differential item function first. A 25-item multiple-choice item test in physics problem solving administered to 650 students flagged the following 'biases':

Table 1

QUEST (Adams and Khoo, 1993) output for compare *routine*

	Adjusted Delta		Difference			
Item #	males d1	females d2	d1-d2	d1-d2 std'ised	Chi-Sq	p
6	-0.94	-0.01	-0.94	-5.11	26.12	0.00
13	0.88	0.46	0.41	2.36	5.57	0.02
14	0.63	1.25	-0.62	-3.45	11.89	0.00

Item No 6: (*easier for males*)

One half-second after starting from rest, a freely falling body will have a velocity of about

A. 2.5 m s^{-1} C. 10 m s^{-1}

B.* 5 m s^{-1} D. 20 m s^{-1}

Items 13 & 14 refer to the following information:

A ball is thrown vertically upwards with an initial velocity of 10 m s^{-1}. Neglect air resistance.

Item No 13: (***easier for females***)

What is the maximum height reached by the ball?

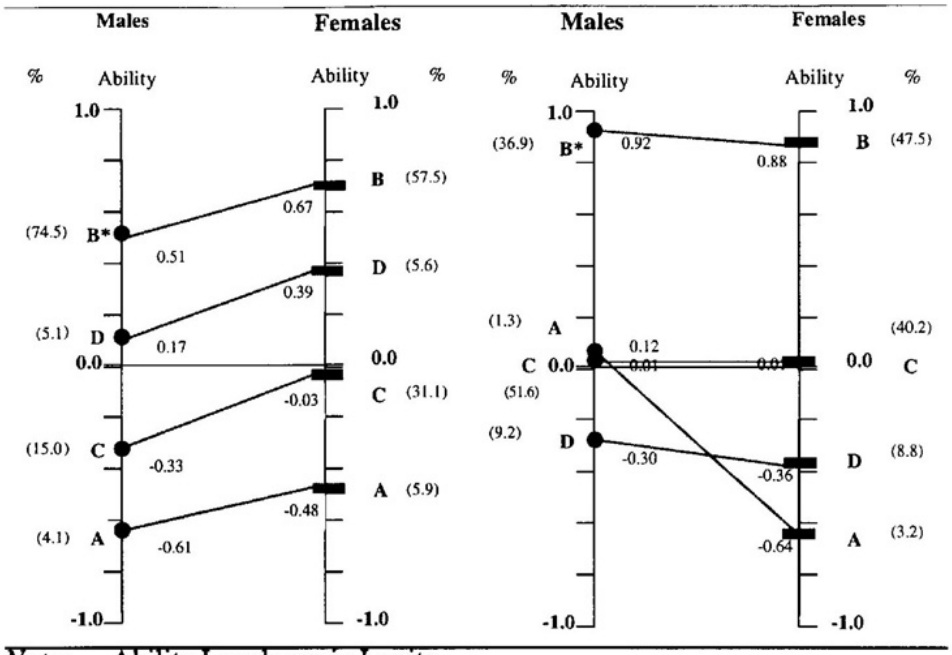
A.	2 m	C.	10 m
B.*	5 m	D.	100 m

Item No 14: (***easier for males***)

How long is the ball in the air?

A.	1 s	C.	5 s
B.*	2 s	D.	10 s

What 'evaded' this generic analyses was that distractors also functioned differentially within Item 13, and was only identified through the Distractor-Ability (D-A) plots devised by Alagumalai & Keeves (1999). The figures below highlight insights gained about this 'teacher constructed' tool.



Note: Ability Levels are in Logits
A, B, C and D are alternatives

Figure 1: D–A Plots for Item 6

Figure 2: D–A Plots for Item 13

In this study, males and females of different 'ability choose distractor A in item 13. This was an important finding, for there may be items that did not show as biased at the item level, but may have differentially function distractors within them. Items 11 and 18, that showed up as neutral, exhibited differential functioning at the distractor levels. The figures below highlight the challenges we may have to examine and identify when constructing a multiple-choice test.

Item No 11: (***neutral***)

At what average power does the person work?

A.	40 W	C.	400 W
B.*	160 W	D.	1600 W

Item No 18: (***neutral***)

A pole AB of length 10 m and weight 800 N has its centre of gravity 4m from the end A, and lies on horizontal ground. The end B is to be lifted by a vertical force applied at B. What is the least force required to do this?

A.	200 N	C.	640 N
B.*	320 N	D.	3200 N

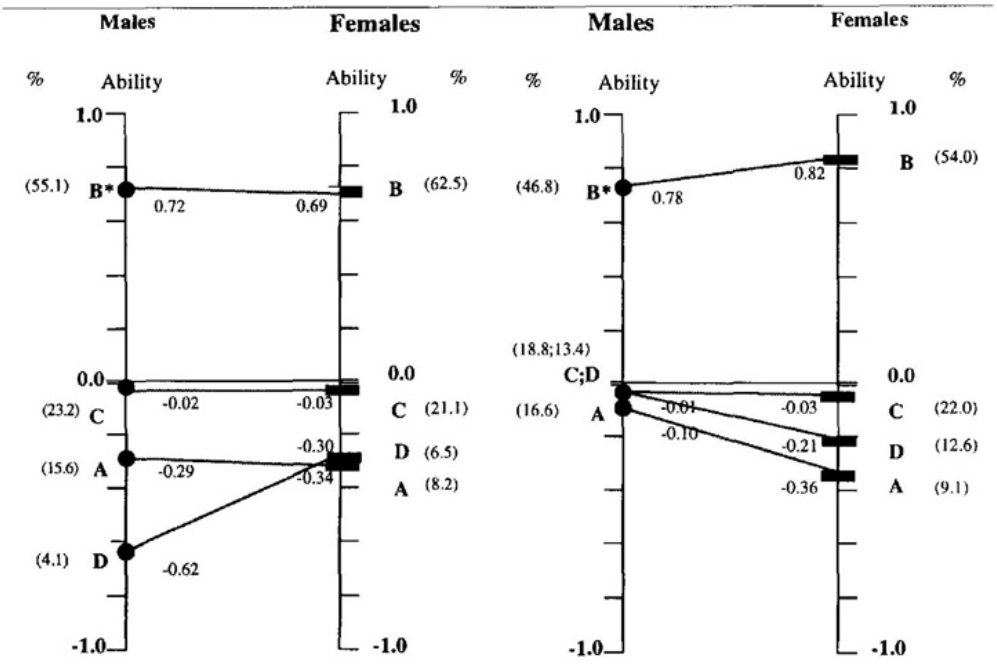


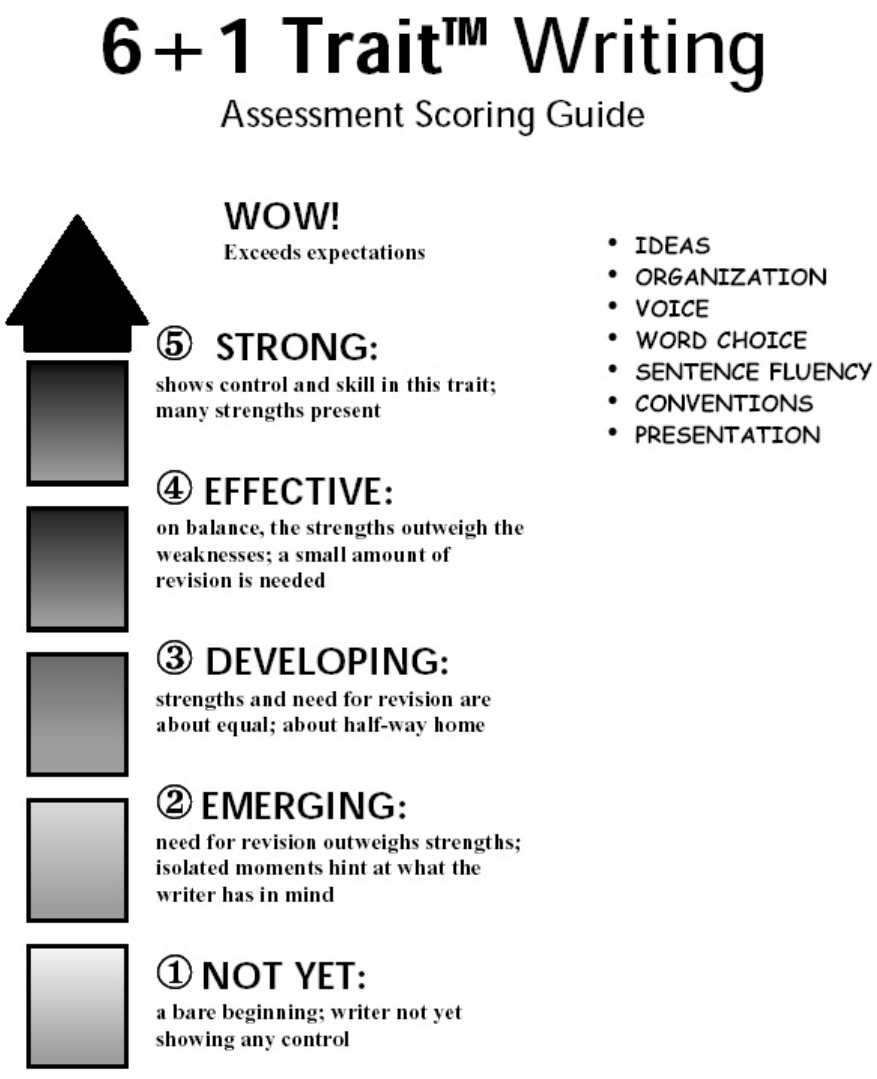
Figure 3: D–A Plots for Item 11

Figure 4: D–A Plots for Item 18

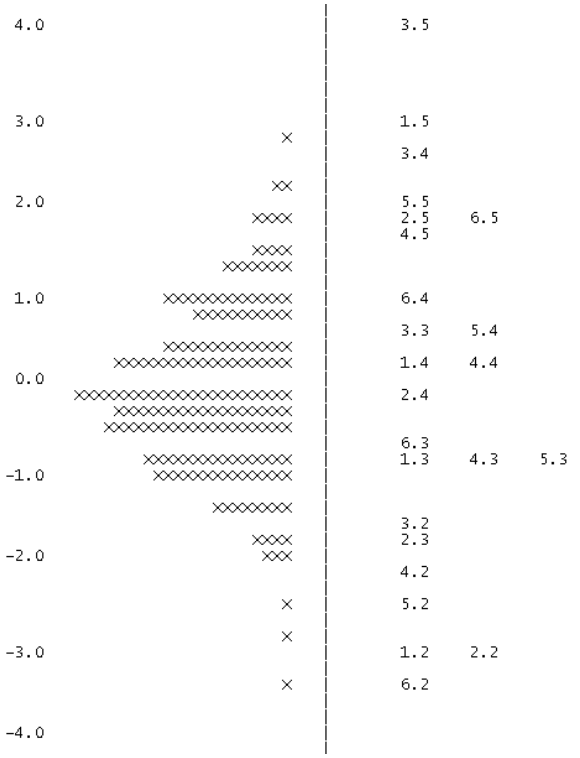
Note: Ability Levels are in Logits
A, B, C and D are alternatives

Case Study Two: Rubrics and Raw Scores (Alagumalai & Sivakumar, 2005; 2006?)

Alagumalai & Sivakumar (2005, 2006?) examined the use of Northwest Regional Educational Laboratory © 6+1 Trait Writing Rubrics (please see screen capture below) in evaluating students essays about the 'Millennium Bug'. Their work examined the perception and understanding of the use of criterion in the 6+1 Writing Rubrics, as well as the assignment of raw scores to students work in the six of the seven categories in the Assessment Scoring Guide.



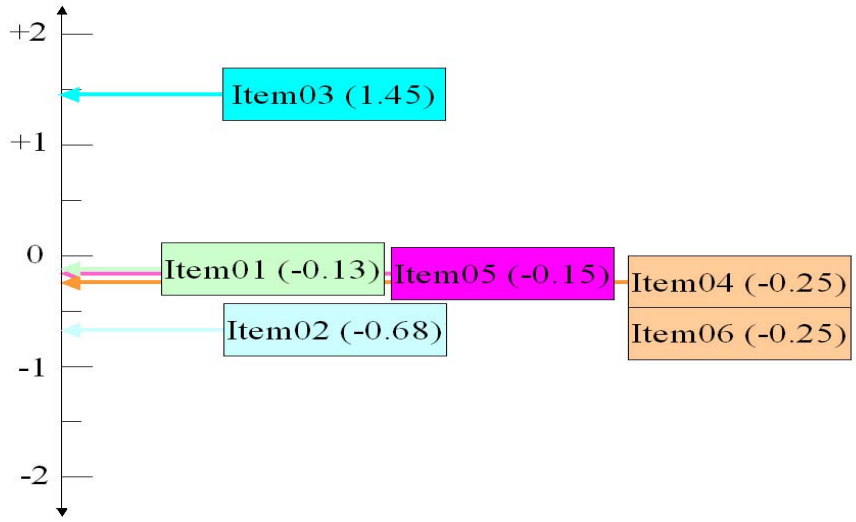
There is an assumption in the 6+1 Writing Rubrics that all categories/items, namely Ideas [1], Organisation [2], Voice [3], Word Choice [4], Sentence Fluency [5] and Conventions [6] all had a maximum score of five, and thus are 'equal' in their difficulty. This may lead to the erroneous assumptions that the raw scores can be aggregated to provide an overall raw score for the Writing Assessment. The Item Analysis using the Quest software highlight that it is very difficult assigning a full score of 5 marks to Word Choice [Category/Item 3] compared with assigning 2 marks for Conventions [Category/Item 6].



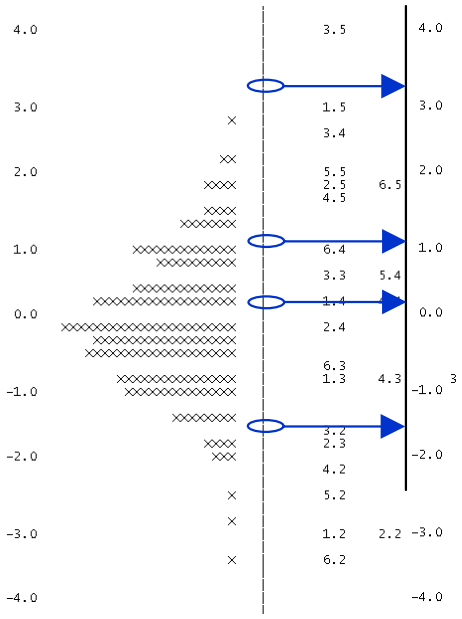

```
Summary of item Estimates
=========================

Mean                              0.00
SD                                0.74
SD (adjusted)                     0.69
Reliability of estimate           0.86
```

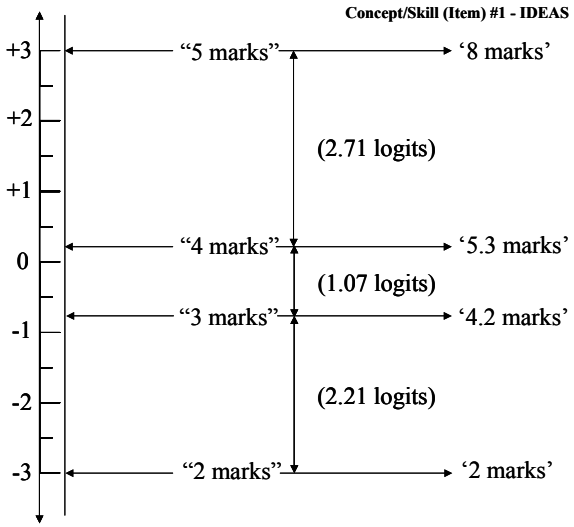
A category/item level plot revels the following pattern:



Hence, Voice as an item/category [3] was relatively more difficult than Organisation [2]. This analysis, using the Rasch model, highlights the fallacy associated with raw scores. The study highlights further the problems with assuming equidistant between the intervals of the raw scores, namely 1 ⇔ 2 ⇔ 3 ⇔ 4 ⇔ 5.



Item-Score Plots for Ideas [1]

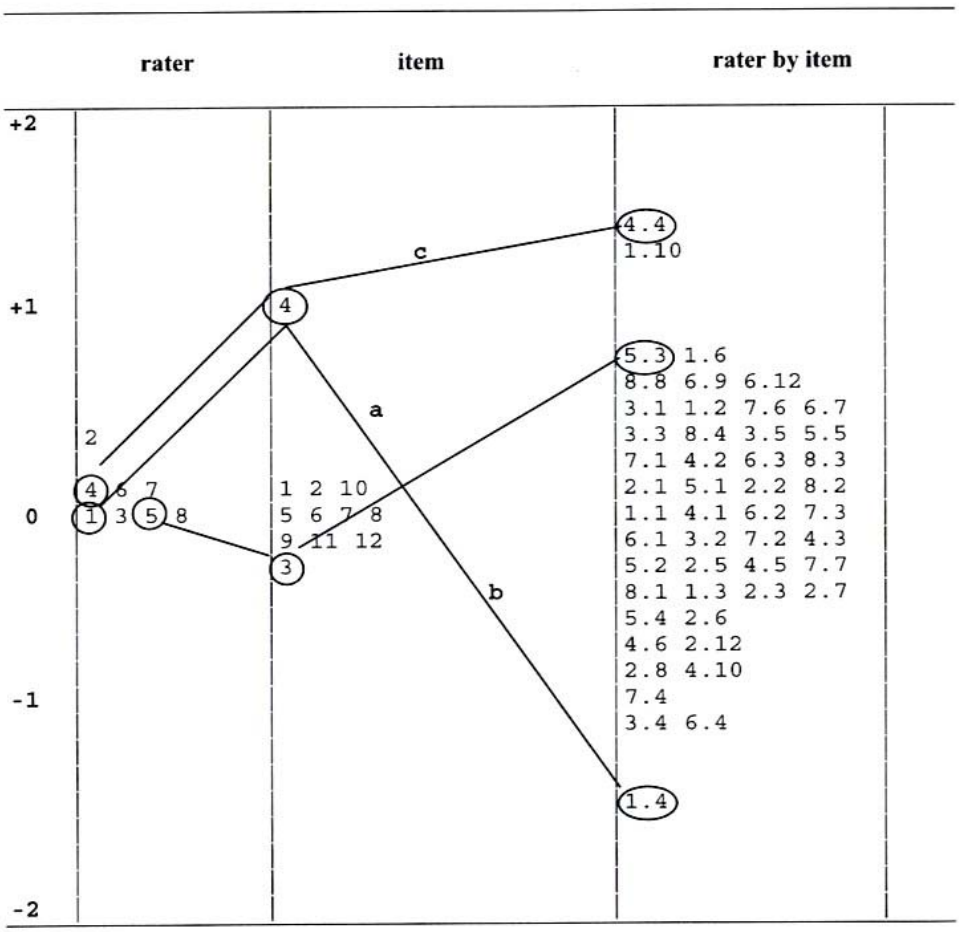


Associated Extrapolation of Scores

Alagumalai & Sivakumar's (2005, 2006?) study highlights the challenges, use and abuse of criterion-based rubrics with rigidly associated raw scores. They argue that post-hoc analyses have to be undertaken to gauge the appropriateness of initial assignment of raw scores.

Case Study Three: Raters' Effects (Barrett 2005, with permission)

Barrett (2005) and Thompson (2004) highlighted the errors associated with raters and judges respectively. Barrett (2005, p.164) identified a number of rater errors, which include 'leniency', 'halo effect', 'central tendency', 'restriction of range', and 'reliability'. In his study, Barrett analyzed the essays of 833 students who had sat for an examination in first year university course in Communication and Media. Eight markers doubled marked 164 scripts, and the figure below represents the Latent Distributions and Response Model Parameter Estimates.



Notes:
Some outliers in the rater by item column have been delted from this figure.
N = 164

Figure 9-5. Map of Latent Distributions and Response Model Parameter Estimates

It is evident that there exists inter- and intra-rater variability. Rater (or marker) [2] is relatively more severe than Rater (or teacher) [1]. It is interesting to note how Rater [5] has found it 'harder' to assign marks to item/question(3), while a relatively less-lenient marker [4] found it 'easier' to score a relatively more difficult item/question (4). Thompson (2004) also provides parallel findings with judging wines, and conclude that judgement against a set of criterion needs scrutiny beyond the face-initial evaluation. Barrett (2005, 176) cautions that a study into rater reliability provides justification for students' concerns. Why isn't one surprised to see shocks in international figure skating and diving competitions?

Discussion and Conclusion

The above exemplary case studies highlight the growing body of knowledge on teacher constructed tools (assessment tasks, tests), and techniques (rubrics to assist in scoring), and subjective judgement. The need for objective measurement and assessment practices sound louder, as articulated by Bond and Caust (2005), especially if many assessments developed by teachers overemphasized memory for procedures and facts (Bransford, 2000, p.245).

Teachers' contribution to students' learning and society is monumental, and need to be supported further by researchers involved in assessment, evaluation and measurement. The partnerships between professional learning and teacher research needs to initiated actively at the pre-service level and continue throughout the professional career.

As argued by de Vaus (cited in Alagumalai, 2005, p.343), "people construct their social world and there are creative aspects to human action, but this freedom will always be constrained by the structures within which people live. Because behaviour is not simply determined we cannot achieve deterministic explanations. However, because behaviour is constrained we can achieve probabilistic explanations." Part of the probabilistic thinking is engaging more actively the broader community of educators, raising fundamental questions of what is the best tool and technique for gauging, assessing (evaluating) and reporting students' learning (Hany, 1998).

Thus, professional development and evidence-based research become the building blocks to address the needs of students in the 21st century and beyond. Hargreaves (cited in MACQT 2005) argues that the process of professionalising teaching must include the following crucial elements:

- Teachers must learn to teach in ways they have not been taught, and who can adjust to changing demands;
- Professional learning must be seen as a continuing process and an individual responsibility as well as an institutional obligation;
- Teachers must have opportunities to learn the skills to become leaders of their colleagues as well as leaders of their classes;
- Teachers must meet an exacting set of professional standards of practice, and be vanguards of educational reforms.

We, as a community of practice, need to interact and collaborate to address the concerns raised by Dunn (2005) about OUR profession. All educators participating in evidence-based research need to address the notion that "Education today is a pre-scientific discipline, reliant upon psychology (philosophy, sociology etc) for its theoretical foundation" (OECD, 2002, p.10). There are growing challenges as articulated through statements like "The science of learning, a branch of human psychology, is still in its infancy. The theory of learning is pre-scientific – in the sense that it lacks as yet either predictive or explanatory power. We do not understand sufficiently well how children and adults learn to dare to offer an educational or training guarantee. The science of education is in its Linnaean phase, drawing up lists of examples of successful learning, clarifying and sorting effective teaching practices; but it still awaits its Darwin with a powerful explanatory theory of learning" (OECD, 2002, p.10). Thus, the assessment design process must be truly multidisciplinary and collaborative activity, with educators, cognitive scientists, subject matter specialists, and psychometricians informing one another during the design process (Pellegrino, 2001, p.314).

There is a growing interdependence between educators (teachers) and researchers in the field of education, and have to take into cognizance characterizing assessments in terms of components of competence and the content and process demands of the subject matter brings specificity to assessment objectives, such as 'higher level thinking' and 'deep understanding'. (Bransford, 2000, p.244). Educators, the public, and particularly parents should not settle for impoverished assessment information. They should be well informed about criteria for meaningful and helpful assessment. To do justice to the students in our schools and to support their learning, we need to recognize that the process of appraising them fairly and effectively requires multiple measures constructed to high standards. Achieving these goals requires a strong connection between educational assessment and modern theories of cognition and learning. (Pellegrino, Chudowsky, & Glaser, 2001, p.314).

In evolving from a highly atomic unidisciplinary curricula through both a interdisciplinary and multidisciplinary curricula and towards a transdisciplinary curricula, not only the pedagogy, epistemology and learning have to be considered seriously, but also the nature of assessment and teacher judgements in the assessment process have to be scrutinized.

While teachers' judgements are scrutinized and refined further, the tools and techniques used by researchers must also be examined. Praxis of action-reflection-action in education coupled with the nexus between teaching-learning-research need to be actualized through collaboration. This means challenging all models advanced in education, and understanding their limitations through critical and reflective practices. As argued by Feynman (1989) "when using a mathematical model, careful attention must be given to the uncertainties in the model." The science of assessment and reporting needs to expand to incorporate diagnostic indices into measurement models to add richer interpretations. (Pellegrino, Chudowsky, & Glaser, 2001, p.137).

To conclude, policy makers have to rethink what constitutes effective and efficient teacher professional development, and how evidence-based research can be synchronized into the daily work of teachers.

References:

Alagumalai, S. (1996). Rasch, SOLO, QUEST and Curriculum Evaluation. Paper presented at the 10th Joint Conference of the Australian Association for Research in Education and the Singapore Educational Research Association. Singapore (25-29 Nov 1996).

Alagumalai, S., & Keeves, J.P. (1999). Distractors – Can they be biased too? Journal of Outcome Measurement, 3(1), pp. 89-102.

Alagumalai, S., Curtis, D.D., & Hungi, N. (2005). Applied Rasch Measurement: A Book of Exemplars. Dordrecht, The Netherlands: Springer.

Alagumalai, S. (2006). Insights into Years 10-12 Physics Problem Solving: A Comparative Study. Dordrecht, The Netherlands: Springer.

Alagumalai, S., & Sivakumar, S. (2006?). Rubrics – Use and Misuse. Journal of Applied Measurement (under review).

Barrett, S. (2005). Raters and examinations. In Alagumalai, S., Curtis, D.D., & Hungi, N. (2005). Applied Rasch Measurement: A Book of Exemplars. Dordrecht, The Netherlands: Springer

Benton, A.L., Sivan, A.B., Hamsher, K.d., Varney, N.L., & Spreen, O. (1994). Contributions to Neuropsychological Assessment. (2nd Ed.) Oxford: Oxford University Press.

Black, P.S. (2004). The Subversive Influence of Formative Assessment. In Alagumalai et al., The Seeker – Reflections and Research. Adelaide, Australia: Flinders University Institute of International Education Publication. [Studies in Comparative and International Education] pp. 77-92

Bond, L. (2004). Teaching to the Test. The Carnegie Foundation for the Advancement of Teaching. Available Online at http://web.uvic.ca/psyc/skelton/Teaching/ [Accessed 21 December 2005]

Bond, T., & Caust, M. (2005). Silk purses from sows' ears? Making measures for teacher judgements. Paper presented at the AARE Conference. Sydney, 2005

Bransford, J.D., Brown, A.L., & Cocking, R.R. ((2000). How People Learn: Brain, Mind, Experience, and School. (Expanded Edition). Washington, D.C.: National Academy Press.

Carlsen, W.S. (1999). Domains of Teacher Knowledge. Cited in Gess-Newsome, J., and Lederman, N.G. Examining Pedagogical Content Knowledge. Dordrecht: Kluwer Academic Publishers. pp. 133-146.

Dunn, S.G. (2005). Philosophical Foundations of Education. Upper Saddle River, NJ: Pearson.

Feynman, R.P. (1989). On the reliability of the Challenger Shuttle. In Feyman, R.P, What do you care what other people think? New York: Bantam

Hany, E.A. (1998). Gifted children in the classroom: Which diagnostic skills do teachers need? Paper presented at the European Council for High Ability Conference. Oxford, UK. 16-19 September 1998.

Harman, A.E. (2001). National Board for Professional Teaching Standards' National Teacher Certification. ERIC Digest. Washington DC: ERIC Clearinghouse on Teaching and Teacher Education (ED460126)

Hassett, M.F. (2000). What Makes a Good Teacher? Cited in Adventures in Assessment, Volume 12 (Winter). Boston, MA: SABES/World Education.

Johnson, V. (2003). Grade Inflation: A Crisis in College Education. NY: Springer-Verlag.

Keeves, J.P., & Alagumalai, S. (1999). New Approaches to Measurement. In Masters, G.N., & Keeves, J.P. Advances in Measurement in Educational Research and Assessment. Amsterdam: Pergamon. pp.23-42.

Keeves, J.P. & Masters, G.N. (1999) Introduction. In Masters, G.N., & Keeves, J.P. Advances in Measurement in Educational Research and Assessment. Amsterdam: Pergamon. pp.1-22.

Kuh, G., & Hu, S. (1999). Unraveling the complexity of the increase in college grades from the mid-1980s to the mid-1990s. Educational Evaluation and Policy Analysis, 21(3), pp.296-320.

MACQT (2005). Towards Greater Professionalisation. Ministerial Advisory Council on the Quality of Teaching Report. (Chapter Four). Available Online at http://www.det.nsw.edu.au/reviews/macqt/macqfi05.htm [Accessed 18 Feb 2006]

OECD (2002). Understanding the Brain: Towards a New Learning Science. Paris, France: Organisation for Economic Co-operation and Development.

Pellegrino, J.W., Chudowsky, N., & Glaser, R. (2001). Knowing what Students Know: The Science and Design of Educational Assessment. Washington, D.C. : National Academy Press.

Pollio, H.R., & Beck, H.P. (2000). When the tail wags the dog: Perceptions of Learning and Grade Orientation in and by Contemporary College Students and Faculty. The Journal of Higher Education, 71(2), pp.84-102.

Thomas, M.D., & Bainbridge, W.L. (1997). Grade Inflation: The Current Fraud. Effective School Research. January Issue.

Thompson, M. (2004). Rasch Scaling and the Judging of Practice. In Alagumalai et al., The Seeker – Reflections and Research. Adelaide, Australia: Flinders University Institute of International Education Publication. [Studies in Comparative and International Education] pp. 145-170

UNESCO (1998). Teachers and Teaching in a Changing World: World Education Report. Place de Fontenoy, 75352 Paris 07 SP: United Nations Educational, Scientific and Cultural Organization

Wink, J. (2005). Critical Pedagogy: Notes from the Real World. (3rd Ed.). Boston: Pearson.

Wright, B.D. (1998). How to convince your friend not to use raw scores. Paper presented at the COMET Meeting. Institute for Objective Measurement & MESA Psychometric Laboratory. 23 Sept 1998.