“The weakest link?”
Enhancing teacher professional judgment in large-scale assessment

Chris Davison,
School of Education, University of New South Wales
Teacher professional judgment is at the cornerstone of much assessment reform in recent years, and has long been the mainstay of school-based assessment systems. However, teacher judgment has always been seen as one of the weakest links in the assessment process with concerns raised about the trustworthiness and reliability of teacher assessment decision-making processes and the teacher's ability to be both “accurate” and “fair”. This presentation will first review some of the concerns and challenges in using teacher professional judgment in large scale school based assessment – both theoretical and practical issues - and then demonstrate how the development of judgment can be supported and enhanced by online professional interaction, including the sharing of student work samples, and benchmarking of assessments. The implications for individual schools as assessment communities will also be discussed.
The background

• Rising international concerns about the perceived inadequacy of teachers’ assessment literacy, defined as a teacher's familiarity with those “measurement basics related directly to what goes on in classrooms”:

"[A]ssessment-literate teachers will typically make better decisions, and because we want students to be better taught, it should be obvious that today's teachers must acquire more assessment literacy" (Popham, 2009, p. 4, 6).
• Concerns have been raised about the trustworthiness and reliability of teacher assessment decision-making processes and the teacher's ability to be both “accurate” and “fair”.

• However, such critics usually treat the (classroom) context as an extraneous variable that must be controlled and neutralized and the (teacher) assessor as someone who must remain objective and uninvolved throughout the whole assessment process (Davison, 2007).

• Four main theoretical problems with these psychometric assumptions ….
1. Assumes standards/criteria can be teacher and context-free

…but assessment criteria always interpreted differently by teachers according to personal background, previous experience, unconscious expectations, internalised and personalised preferences regarding the relative importance of different criteria and ideological orientation

"We are social beings who construe the world according to our values and perceptions"

*(Gipps 1994: 167)*
2. Assumes assessment is essentially a technical activity

…but assessment criteria can never be made explicit, indicators are always ambiguous, and require the application of implicit knowledge (Wiliam 2001, Claxton 1995)

e.g teachers only come to understand and interpret criteria by looking at actual student texts (and contexts) and talking with each other, not only about the impact on the student, but the impact on learning and teaching
3. Assumes teachers will accept externally imposed criteria and assessments as basis for judgment …but growing evidence that when assessments conflict with teachers’ own personalised judgments teachers manipulate and/or reject criteria (Davison, 1999, Arkoudis & O’Loughlin 2002)
4. Assumes agreement (with criteria) rather than disagreement essential for ‘validity’ and ‘reliability’

…but even if it is possible to define criteria unambiguously, not necessarily desirable (Mabry 1999, Smith, 1991)

eg. “criterion-referenced hyperspecification” (Popham 1994) can lead to students and teachers predicting too accurately what is to be assessed, over-emphasising and/or learning/teaching only aspects of the curriculum that are (able) to be assessed, thus compartmentalising learning/teaching

“Put crudely, the more precisely we specify what we want, the more likely we are to get it, but the less likely it is to mean anything” (Wiliam 2001)
• Teacher-based assessment derives a major part of its validity and reliability from its location in the actual classroom where assessment activities are embedded in the regular curriculum, and assessed by a teacher who is familiar with the student’s work and has – presumably – a stake in their improvement.

• However, schools and teachers need to be granted a high degree of trust and autonomy in design, implementation and timing of assessment tasks.

• Students also need to be given sufficient time and support to demonstrate their best - to show what they can do - and for the assessor to be able to confidently assess their output, but even more importantly, validate their informal judgments of students’ language levels and achievements.
• More formal assessment tasks should encourage the teacher to stand back and reflect on their implicit or explicit assumptions about individual students’ capacities, compare those assumptions with careful analysis of examples of students’ actual performance, and then subject their judgments to explicit scrutiny and challenge, or confirmation by others.

• In other words, we do not assume that the class teacher is objective and/or has no preconceived ideas or assumptions about a student’s level. Rather, we seek to make such assumptions explicit and open to discussion with fellow teachers.

• Trustworthiness comes more from the process of expressing disagreements, justifying opinions and validating them than from immediate consensus.
• Even when it is possible to establish common understandings of the task, publicly agreed and explicit assessment criteria, and strong moderation processes, there will always be the problem of interpretation. Hence, professional dialogue and interaction central to decision-making process. These are inherent strengths, not weaknesses, of teacher based assessment systems.

• However, “in order to maintain trust communities will have to show that their procedures for making judgements are fair, appropriate and defensible (i.e. that they are valid), even if they cannot be made totally transparent” (Wilam 2001)
The way forward ....

1. Assume teachers are professionals, not technicians
2. Assume learner, text, context and community all important and inter-related factors in making assessment decisions
3. Assume assessment is embedded in classroom practice, but classrooms are not isolated nor autonomous units, but part of a community of assessment practice
4. Assume assessment decisions grounded in complexity and conflict, conflict accepted, even embraced, as a continuing condition which will lead to greater understanding.
The practical problems?

• Look at these video excerpts of “teacher assessment” performance:

  How effectively can these teachers assess their students?

• What knowledge, skills and attitudes do they draw on to make their decisions?
• What tools and resources helped/would have helped them in their decision-making?
School-Based Assessment in Hong Kong
Introduction to School-based Assessment

SBA in the 2012 HKDSE English Examination

Assessment, Teaching and Learning: from Principles to Practice

Extensive Reading Programme

Case Studies: Sample Teaching, Learning and Assessment Sequences

Appendices

Acknowledgements
• High quality assessment decision–making depends on:
  – Our knowledge and assumptions about what is involved in assessment decision-making (How trusting we are?)
  – Our access to trustworthy tools for assessment (eg. criteria, exemplars), benchmarking and self and peer reflection (How trustworthy we are?)
  – The role and status of our assessment decision making: (How trusted are we?)
Recent relevant research:
Davison, Michell & Chee (2012a,b)

– An examination of teacher-based practice in assessment (including relevant aspects of existing ESL/EAL assessment practices) in Victoria and NSW

– 30+ experienced EAL specialist teachers from all three systems and levels of schooling examined the suitability of a selected range of assessment tools for their teaching context, and discussed what they wanted/needed for effective assessment of EAL development
Example: Victorian and NSW teachers …

• How trusting are we?

“I find when I’m doing a multiple choice type of test with my kids that, many of them, just guess and I watch them … because they can’t understand really what’s asked of them. … and that gives a really false - yes, they may be right, but you’re never really sure of how much they know and how much it’s all simple guesswork”
(110506_09016.40-7.25)
“I teach grade 1/2, … and some of my children got 100%, so that the test wasn’t any good for them, but some of them couldn’t get any right and they’re all on the same grade, so if I were just to give them that test and they were sent away, it’s not going to give me any information. I already know who might … which children will do well and which wouldn’t. I need more information than that.” (110505_1130 36.49-37.16)
• How trustworthy are we?

“From my observation, you can give five teachers of this school a piece of writing and get them to place it on the VELLS continuum and [get] five different responses – it’s so subjective, so any test that would help us ... across the board, that would be useful.”

(110506_0333 28.29-29.02)”
“To have a reliable assessment tool, that we will actually assess the child and provide you with that information, I think, that is something that we definitely need, and is lacking. And for me, as an ESL teacher... I’m really longing for something like that to be around because it’s otherwise a bit of a guessing game ... and, I think, you know, we’re not, we’re not really supported by data, as yet.” (110505_1315 22.02-23.28)
“(It) would be valuable to have some consistent approaches so we have a sort of common language, or common tasks when we’re assessing, particularly in terms of transitions when students leave our language school and they go to a mainstream school, or to an independent school, um, we have, yeah, a clear sense of where they’ve come from and the language that’s been used by the teacher in assessing them, reporting their progress.” (01.01.19-01.01.40)
• How trusted are we?

“I was really upset because it’s completely inappropriate, it gives me information that I am already able to gather in my own classroom ... the texts are culturally exclusive ... the experience is quite traumatic for my students, the child just sits there, and there is a sense of feeling of failure”
Findings:

• In the absence of systematic standardized approaches to EAL assessment, EAL teachers’ development and/or use of classroom- or school-based ESL assessments, was found to be limited, localised and ad hoc, and often not valued or recognised as “assessment”.

• Teachers’ knowledge about assessment was mainly informal, and professional learning tended to be confined to informal trial-and-error implementation of classroom assessment strategies.

• Thus, clear need for more assessment literacy, but also more effective standardised assessment instruments that could be tailored for the teacher’s context, and stronger (online) assessment learning communities.
“The ideal of being able to create your own tests was something that appealed to us. And .... different types of reports that could be generated to inform the teacher about what the child knows ... to be able to identify what the student can do, what the student needs to do ... and then the next step, you know, there are ideas for ... the actual teaching of those particular areas” (110505_1130 14.18-15.11)

“The best kind of testing is what you generate yourself and that relates to what you're doing in the classroom” (210505_360 11.28-11.29)

“I was thinking more an evaluation package rather than a specific tool. It would be really good to have some type of (coherent) resource (to help student learning)” (110505_1130 37.38-38.07)
In NSW …

“The new Years 7-10 syllabuses advocate assessment for learning, as this is a type of quality assessment that has had world-wide success in enhancing teaching and improving student learning. Assessment for learning gives students opportunities to produce work that leads to development of their knowledge, understanding and skills. Teachers decide how and when to assess student achievement, as they plan the work students will do, using a range of appropriate assessment strategies including self-assessment and peer assessment.”

http://arc.boardofstudies.nsw.edu.au/go/7-8/assessment-for-learning-in-years-7-10/

“Assessment for learning:

– emphasises the interactions between learning and manageable assessment strategies that promote learning

– clearly expresses for the student and teacher the goals of the learning activity

– reflects a view of learning in which assessment helps students learn better, not just achieve a better mark

– provides ways for students to use feedback from assessment

– helps students take responsibility for their own learning

– is inclusive of all learners”
Key characteristics of ‘assessment” in AfL:

a. assessment is embedded in teaching and learning;
b. learning goals are explicitly shared with students and students are taught how to know and to recognize the standards they are aiming for;
c. students are engaged in continuous peer and self-assessment;
d. constructive qualitative feedback helps students to recognize the next steps needed for learning and how to take them;
e. teachers, parents and students regularly review and reflect on assessment data;
f. it is assumed every student can improve;

(adapted from the Assessment Reform Group, 1999, p.7)
Why? Strong and sustained research base...

- Black & Wiliam (1998) have convincingly demonstrated the learning gains that can be achieved through well focused teacher-based formative assessment enhance students’ learning more than any other strategy across age levels and in different contexts.

- Hattie (2009) in a more recent study of major influences on educational achievement (using 800+ meta-analyses) found that formative practice, in particular self-assessment and feedback, had the highest effect sizes (i.e., impact on student outcomes) out of more than 100 different instructional and contextual factors.
Establishing the research questions

Significance of the study

Literature search

Workshop with teachers and principals

Experts’ validation

Pilot testing

Exploratory factor analysis

Confirmatory factor analysis

Parametisation (Item response theory application)

Rationale for choosing the construct

Defining the construct

Indicators

Performance criteria

AfL competency performance instrument

Item characteristics

Hierarchy of AfL indicators

Stage 1
• Validation by 60 EAL Teachers in NSW

Stage 2
• Validation by 32 EAL Teachers in Victoria

Stage 3
• Tested with 98 EAL Teachers/ Analysis

The findings of the EAL Teachers’ Assessment for Learning Competency Framework (Alonzo & Davison, 2014)
Teacher self-perceptions of assessment literacy

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Mean Max = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conducts assessment with consideration of student background and culture</td>
<td>4.36</td>
</tr>
<tr>
<td>Participates in professional development related to assessment</td>
<td>4.36</td>
</tr>
<tr>
<td>Demonstrates belief in the ability of every student to improve</td>
<td>4.14</td>
</tr>
<tr>
<td>Identifies appropriate teaching methods</td>
<td>3.94</td>
</tr>
<tr>
<td>Maintains confidentiality in dealing with assessment results (norm referencing, inappropriate disclosure of assessment results and ranking students)</td>
<td>3.93</td>
</tr>
<tr>
<td>Develops an environment of trust</td>
<td>3.93</td>
</tr>
<tr>
<td>Undertakes further education/ training in assessment</td>
<td>3.93</td>
</tr>
<tr>
<td>Tailors lessons to available resources</td>
<td>3.88</td>
</tr>
<tr>
<td>Uses assessment to build students’ interest to learn</td>
<td>3.86</td>
</tr>
<tr>
<td>Reinforces positive learning attitude of students</td>
<td>3.79</td>
</tr>
</tbody>
</table>
## Teacher self-perceptions of assessment literacy

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gathers a range of evidence of student learning</td>
<td>2.36</td>
</tr>
<tr>
<td>Designs English language assessment tasks</td>
<td>2.45</td>
</tr>
<tr>
<td>Engages students in peer-assessment</td>
<td>2.45</td>
</tr>
<tr>
<td>Engages in self-assessment/ reflection</td>
<td>2.68</td>
</tr>
<tr>
<td>Involves students in the development of learning outcomes</td>
<td>2.87</td>
</tr>
<tr>
<td>Gives feedback related to criteria</td>
<td>2.87</td>
</tr>
<tr>
<td>Assists students in using feedback to feed forward</td>
<td>2.98</td>
</tr>
<tr>
<td>Collaborates with family to establish home activities to support students</td>
<td>2.98</td>
</tr>
<tr>
<td>Informs community of school’s assessment practices</td>
<td>3.12</td>
</tr>
<tr>
<td>Develops appropriate English language assessment strategies</td>
<td>3.33</td>
</tr>
<tr>
<td>Moderates feedback and results of self and peer assessment</td>
<td>3.33</td>
</tr>
<tr>
<td>Engages in peer-review of teaching performance</td>
<td>3.36</td>
</tr>
<tr>
<td>Identifies key assessment and teaching issues for review</td>
<td>3.36</td>
</tr>
<tr>
<td>Involves students in the development of success criteria/ rubrics</td>
<td>3.4</td>
</tr>
<tr>
<td>Explains the success criteria/ rubrics</td>
<td>3.4</td>
</tr>
<tr>
<td>Engages students in self-assessment</td>
<td>3.4</td>
</tr>
</tbody>
</table>
Tools to enhance assessment literacy for teachers of EAL (2012-2015)

- Drawing on Assessment for Learning (AfL) principles and Vygotskian theory, the TEAL project aims to develop and validate an online assessment advice and ‘toolkit’, http://teal.global2.vic.edu.au/ (under development) for use by all teachers to provide a measure of the stage of development for EAL students in speaking and listening, reading and writing (and highlight the implications for teaching)

- All tools to be aligned against the Victorian ESL Standards (VELS) and EAL Developmental Continuum, with potential for alignment to other standards by other jurisdictions.
Welcome to Tools to Enhance Assessment Literacy

The TEAL web project, recently launched in 2014, is an online resource for teachers of primary and secondary level children who are learning English as a second (ESL) or additional (EAL) language in Australia. It brings together a range of tools and advice for the assessment and reporting of the English language proficiency and progress of students.
Project outcomes (1)

• An online ESL Assessment Resource Centre
  • Teacher professional learning resources - containing background material re assessment principles and processes, plus with video and text-based resources.
  • An assessment tools bank – containing a range of assessment tools and tasks organised around listening & speaking, reading and writing cross-referenced by assessment type (observation, test, analysis etc), VELS ESL stages and Year levels.
• **Assessment for teaching and learning exemplars** –
  containing a selection of annotated units of work across a range of subject areas and year levels showing assessment tasks with formative feedback embedded within a teaching/learning cycle.

• **An online teacher discussion forum** - password-protected area for teachers to share problems, strategies and work samples
Project outcomes (2)

- A prototype teacher-based assessment system
  - for the collection and analysis of oral and written language samples and exemplars aligned with the ESL Continuum to provide information on students’ English language and literacy development.
  - includes strategies for evaluating students’ L1 language and literacy development.
A prototype teacher-based assessment system for the collection and analysis of oral and written language samples and exemplars aligned with the ESL Continuum to provide information on students' English language and literacy development. Includes strategies for evaluating students' L1 language and literacy development.
A rabzé is half rabbit and zebra and cat. There was a rabzé banana and ant, and frog and worm deer monkey in garden in tree jungle. Rabzé a little one. Have 100 rabzé. Have little, baby rabzé.
Uniswan

This is my own animal. Her name is Uniswan. She is a unicorn with a swan wings and tail. She can fly too! She live in a wet place. Her skin is shiny, and also, her tail is very hard, because when she see enemies, she will use her tail to touch her enemies.

She very like to help. She can make rainbow. When the people sick or hurt, she'll use her rainbow to help.
Task 6 ‘An imaginary animal’  Sample 3

Camos

A camgo is half camil and half kangaroo. It is blue in colour with pink strips. Camgos likes to eat leaves and if they see people they will run as fast as a car and eat them too. You can find Camgos in the school roofs and in your car bubs and did you now camgos can be in your lunchboxes too because they are little and some are big. Camgos likes to play around with their friends and family. But you have to be careful because if they are angry they will chase they make troubles and kill them. So check your car before you go and hang around with your friends.
Teacher feedback  Task 6 ‘An imaginary animal’

**Grade 5 and 6, Stages BL, B1, B2**

- ‘students enjoyed the lesson’
- ‘used a mixture of imaginary animals and real animals’
- ‘able to be for low and high needs students (eg. new arrivals)’
- ‘drawing of animals was great as away of beginning - all students felt success’
- ‘students could write as little or as much as they were able’

**Grade 4 and 6, Stages BL, B3**

- ‘students thoroughly enjoyed writing about their imaginary animal’
- ‘they liked the idea of creating their own animal and describing it’
- ‘there were no set boundaries so they felt at ease being able to write and describe their very own creations’
- ‘they developed a sense of pride taking ownership for their work’
- ‘students enjoyed the mix and match game to create new animals’

**Grade 2, Stage B2**

- ‘highly engaging task , especially in the tuning in part of the lesson... prompted motivation for writing’
- ‘students could relate due to prior experiences and knowledge of this type of genre’
- ‘students were able to use common adjectives to describe their imaginary animal’
- ‘they developed a sense of pride taking ownership for their work’
- ‘students relied on books/websites (in simple language) to find information about their preferred animal and to support heir writing’

NOTE: Task did not rely on specific subject content, but at the same time samples the academic knowledge and skills required to be successful with that content.
Project outcomes (3)

• A calibrated item bank of reading and vocabulary items linked to texts
  • to be used in a computer adaptive testing (CAT) system aligned with ESL continuum to provide information on students’ English and L1 language and literacy development at the four NAPLAN assessment points Years 3, 5, 7 and 9.
  • being undertaken with Educational Assessment Australia, UNSW Global.
ESL teacher participation

• Within each stage, the project has drawn on the specific professional knowledge of EAL teachers in:
  • collecting, evaluating and developing exemplar school-based assessment materials, tasks and strategies,
  • writing and critical review of assessment tasks items, and in
  • providing feedback on existing and recommended assessment practice.
The theoretical framework

Contingent scaffolding (just-in-time support)
- hints, prompts, questions
- modeling, demonstration
- linking, weaving
- discursive pressure
- gesturing, positioning

‘Designed in’ scaffolding
- mediating tools, texts, videos
- meta-language
- procedural/conceptual/social activity routines and sequences

Means/tools/instruments

Assessment activity system

Agent
- teacher as pedagogue

Object
- student task participation & mastery of task operations

Rules
- task ownership
- task challenge
- task completion
- success criteria

Community
- classroom
- groups
- dyads

Division of labour
- helper/helpee
- expert/novice
- teacher/student
- handover/take-over

Outcome
- learner/teacher independence and confidence
- agency ‘spin-offs’
- transfer of skills to other (like ?) tasks

School of Education
Scaffolding is the timely, temporary and partial assistance provided by experts to novices to enable their successful participation or performance on new and difficult intellectual tasks.
The theoretical framework

- **controlled**
- **scaffolding**
- **guided**
- **independent**

- **other regulated activity**
- **self regulated activity**

- **participation**

- **Students/teachers**

- **Teacher/TEAL**

- **time**

School of Education
A dialectic model of teacher assessment judgement (Michell, in preparation)

- **global judgement**
  - mediating tools
    - constructs
    - standards
    - relations
    - community
  - cognition
    - tacit knowledge
    - intuition
    - ‘on balance’ judgement
  - analytical judgement
    - text
    - rubrics
    - criteria
    - procedures
    - explicit knowledge
    - logic
Conclusions

This project demonstrates that with support and collaboration and support we can build an online assessment “system” which

- puts the learners and teachers at the centre of the assessment process
- is theoretically and philosophically coherent
- can model desired outcomes and scaffold and support sustainable improvements in assessment, learning and teaching
- can build an assessment–literate community of not just teachers, but students, parents and other key stakeholders
- can lead to differentiated contextualised tailor-made assessment practices that also have commonality, consistency and public confidence