Hawker Brownlow EDUCATION

12th Annual

# Thinking & Learning Conference

## DR DYLAN WILIAM

Friday 22 May

## Formative Assessment – What it is and What it isn't

*Session 1*

## MELBOURNE


Hawker Brownlow EDUCATION

03 8558 2444

03 8558 2400

www.hbconf.com.au

conferences@hbe.com.au

# DYLAN WILIAM

Dylan (PhD) is Emeritus Professor of Educational Assessment at University College London. In a varied career, he has taught in urban public schools, directed a large-scale testing program, served a number of roles in university administration and pursued a research programme focused on supporting teachers to develop their use of assessment in support of learning.

## A message from Hawker Brownlow Education

We hope that you have found these conference papers and the accompanying sessions useful. Please be aware that the contents of these papers are the intellectual property of the speaker and no reproduction for any purpose is authorised. We urge you to take care of this booklet. Replacement copies will not be made available either during or after this conference.

Published in Australia by

Hawker Brownlow EDUCATION



Hawker Brownlow Education
P.O. Box 580, Moorabbin, Victoria 3189, Australia
Phone: (03) 8558 2444 Fax: (03) 8558 2400
Toll Free Ph: 1800 334 603 Fax: 1800 150 445
Website: www.hbe.com.au
Email: orders@hbe.com.au


Printed in Australia

CODE: MELDW0101
0515

# Formative assessment: what it is and what it isn't; when it works and when it doesn't

Dylan Wiliam (@dylanwiliam)

www.dylanwiliam.net

## Formative assessment as *assessment*

2

- Assessment is a procedure for making inferences (Cronbach, 1971)
- Four elements of educational assessment
  - Designing ways in which we can get evidence relevant to student learning
  - Collecting the evidence
  - Interpreting the evidence
  - Using the evidence

## Formative and summative assessment

3

- The formative-summative distinction is a classification of the kinds of inferences being made
  - Summative: inferences regarding an individual's current, or future, status
  - Formative: inferences regarding the kinds of instructional activities likely to improve learning
- No such thing as *a* formative or *a* summative assessment

## Issues in defining formative assessment

4

- Theorization and definition
  - Possible variables
    - Category (instruments, outcomes, functions)
    - Beneficiaries (teachers, learners)
    - Timescale (months, weeks, days, hours, minutes)
    - Consequences (outcomes, instruction, decisions)
    - Theory of action (what gets *formed*?)

## Which of these are formative?

5

A. A district science supervisor uses test results to plan professional development workshops for teachers
B. Teachers doing item-by-item analysis of 5th grade math tests to review their 5th grade curriculum
C. A school tests students every 10 weeks to predict which students are "on course" to pass the state test in March
D. "Three-fourths of the way through a unit" test
E. Students who fail a test on Friday have to come back on Saturday
F. Exit pass question: "What is the difference between mass and weight?"
G. "Sketch the graph of y equals one over one plus x squared on your mini-white boards."

## Issues in defining formative assessment

6

- Theorization and definition
  - Possible variables
    - Category (instruments, outcomes, functions)
    - Beneficiaries (teachers, learners)
    - Timescale (months, weeks, days, hours, minutes)
    - Consequences (outcomes, instruction, decisions)
    - Theory of action (what gets *formed*?)

## An inclusive definition

7

An assessment functions formatively:

"to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited." (Black & Wiliam, 2009 p. 9)

## Consequences of the definition

8

- Formative assessment is not a thing
- Anyone—teacher, peer, learner, can be the agent of formative assessment;
- The focus is on decisions, rather than data;
- The next steps in instruction indicated may not be the best, or even successful;
- The assessment need not actually change the decision made.
- And therefore, formative does not mean optimal, or even good

## Formative assessment: generic *and* specific

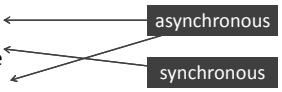
9

- Formative assessment has both generic and subject specific aspects
- The five strategies of formative assessment are generic, but they play out differently in different subjects
- There is a continuing debate about what aspects of formative assessment are generic (pedagogy) and which are domain-specific (didactics)

## FA and the regulation of learning processes

10

- Formative assessment is therefore concerned with the creation of, and capitalization upon, moments of contingency in the regulation of learning processes
- Kinds of regulation (Allal, 1988)
  - proactive
  - interactive
  - retroactive




## The formative assessment hijack

11

- Long-cycle:
  - Span: across units, terms
  - Length: four weeks to one year
  - Impact: Student monitoring; curriculum alignment
- Medium-cycle:
  - Span: within and between teaching units
  - Length: one to four weeks
  - Impact: Improved, student-involved assessment; teacher cognition about learning
- Short-cycle:
  - Span: within and between lessons
  - Length:
    - day-by-day: 24 to 48 hours
    - minute-by-minute: five seconds to two hours
  - Impact: classroom practice; student engagement

## Main approaches to formative assessment

12

- Professional Learning Communities

  "…an inclusive group of people, motivated by a shared learning vision, who support and work with each other, finding ways, inside and outside their immediate community, to enquire on their practice and together learn new and better approaches that will enhance all pupils' learning." (Stoll et al., 2006)
- Two main approaches
  - Focus on outcomes for students
  - Focus on increased teacher capacity

## Complementary processes

13

| Data-driven PLCs | Classroom FA TLCs |
|---|---|
| • Quality control<br>• Common assessments<br>• Improvement through better team work and systems<br>• Focus on individual outcomes for students<br>• Regular meetings focused on data<br>• 16 points on PISA (in two to three years) | • Quality assurance<br>• Highly structured meetings<br>• Improvement through increased teacher capacity<br>• Focus on teachers' individual accountability for change<br>• Regular meetings focused on teacher change<br>• 30 points on PISA (in two to three years) |

## Unpacking formative assessment

14

| | Where the learner is going | Where the learner is | How to get there |
|---|---|---|---|
| Teacher | Clarifying, sharing and understanding learning intentions | Engineering effective discussions, tasks, and activities that elicit evidence of learning | Providing feedback that moves learners forward |
| Peer | | Activating students as learning resources for one another | |
| Learner | | Activating students as owners of their own learning | |

## Formative assessment and other priorities

15

- Formative assessment is an integral part of many current policy priorities:
  - Framework for teaching (Danielson)
  - Differentiated instruction (Tomlinson)
  - Response to (instruction and) intervention

## FA has included SRL for years

16

"Formative assessment is concerned with how judgments about the quality of student responses (performances, pieces, or works) can be used to shape and improve the student's competence by short-circuiting the randomness and inefficiency of trial-and-error learning." (Sadler 1989 p. 120)

"The indispensable conditions for improvement are that the student comes to hold a concept of quality roughly similar to that held by the teacher, is able to monitor continuously the quality of what is being produced during the act of production itself, and has a repertoire of alternative moves or strategies from which to draw at any given point. In other words, students have to be able to judge the quality of what they are producing and be able to regulate what they are doing during the doing of it." (p. 121)

## Motivation, self-regulation and learning

| | | Value system | |
|---|---|---|---|
| | | External | Internal |
| **Locus of control** | External | External regulation | Identified regulation |
| | Internal | Introjected regulation | Integrated regulation |

Deci and Ryan (1994)

## Framework for teaching (Danielson 1996)

18

- Four domains of professional practice
  1. Planning and preparation
  2. Classroom environment
  3. Instruction
  4. Professional responsibilities
- Links with student achievement (Sartain, et al. 2011)
  - Domains 1 and 4: no impact on student achievement
  - Domains 2 and 3: some impact on student achievement

## The framework in detail

19

- Domain 2: The classroom environment
  - 2a: Creating an environment of respect and rapport
  - 2b: Establishing a culture for learning
  - 2c: Managing classroom procedures
  - 2d: Managing student behavior
  - 2e: Organizing physical space
- Domain 3: Instruction
  - 3a: Communicating with students
  - 3b: Using questioning and discussion techniques
  - 3c: Engaging students in learning
  - 3d: Using assessment in instruction
  - 3e: Demonstrating flexibility and responsiveness

## Formative assessment and domain 3

20

| Framework for teaching | Classroom formative assessment |
|---|---|
| Communicating with students | Sharing learning intentions with students |
| Using questioning and discussion techniques | Eliciting evidence |
| Engaging students in learning | Feedback |
| Using assessment in instruction | Students as learning resources |
| Demonstrating flexibility and responsiveness | Students as owners of their learning |

## Differentiated instruction: not a new idea

21

- Differentiation in action (Stradling & Saunders, 1993)
- Differences in
  - educational goals
  - curriculum structure
  - course content
  - learning tasks
  - teaching approach
  - pace of learning
  - assessment
  - review

## Most definitions of DI are vague

22

"While the concept of 'differentiated instruction' can be defined in many ways, as good a definition as any is ensuring that what a student learns, how he/she learns it, and how the student demonstrates what he/she has learned is a match for that student's readiness level, interests, and preferred mode of learning." (Tomlinson, 2004 p. 188)

"To differentiate instruction is to recognize students' varying background knowledge, readiness, language, preferences in learning and interests; and to react responsively. Differentiated instruction is a process to teaching and learning for students of differing abilities in the same class." (Hall, Strangman, & Meyer, 2011)

## Differentiated instruction and formative assessment

23

| | Aspects of differentiated instruction (Hall, Strangman & Meyer, 2008) | FA? |
|---|---|---|
| Content | Several elements and materials are used | |
| | Align tasks and objectives to learning goals | |
| | Instruction is concept-focused and principle-driven | |
| Process | Flexible grouping is consistently used | |
| | Classroom management benefits students and teachers | |
| Products | Initial and on-going assessment of student readiness and growth | |
| | Students are active and responsible explorers | |
| | Vary expectations and requirements for student responses | |
| Miscellaneous | Clarify key concepts and generalizations | |
| | Use assessment as a teaching tool | |
| | Emphasize critical and creative thinking as a goal in lesson design | |
| | Engaging all learners is essential | |
| | Balance between teacher-assigned and student-selected tasks | |

## Response to (instruction and) intervention

24

"Response to intervention integrates assessment and intervention within a multi-level prevention system to maximize student achievement and reduce behavior problems. With RTI, schools identify students at risk for poor learning outcomes, monitor student progress, provide evidence-based interventions and adjust the intensity and nature of those interventions depending on a student's responsiveness, and identify students with learning disabilities." (National Center on Response to Intervention, 2010)

- Two "creation myths" for RT(I)I
  - A protocol for preventing academic failure (progress monitoring, early—research-based—intervention)
  - An alternative to IQ testing in the identification of learning disabilities

## Response to (instruction and) intervention

25

- Key points
  - Tier 1 must be high-quality, evidence-based instruction
  - Student progress must be monitored
  - Failure to progress triggers additional support
- Formative assessment
  - Makes tier 1 instruction as effective as it can be
  - Allows assessment of progress (for tier 2 assessment)

## And one big idea

26

| | Where the learner is going | Where the learner is | How to get there |
|---|---|---|---|
| Teacher | | | |
| Peer | | | |
| Learner | | | |

Using evidence of achievement to adapt what happens in classrooms to meet learner needs

## An educational positioning system

27

- A good teacher:
  - Establishes where the students are in their learning
  - Identifies the learning destination
  - Carefully plans a route
  - Begins the learning journey
  - Makes regular checks on progress on the way
  - Makes adjustments to the course as conditions dictate

## Force-field analysis (Lewin, 1954)

28

| + | — |
|---|---|
| What are the forces that will support or drive the adoption of formative assessment practices in your school/district? | What are the forces that will constrain or prevent the adoption of formative assessment practices in your school/district? |

## The evidence base for formative assessment

29

- Fuchs & Fuchs (1986)
- Natriello (1987)
- Crooks (1988)
- Bangert-Drowns, et al. (1991)
- Dempster (1991, 1992)
- Elshout-Mohr (1994)
- Kluger & DeNisi (1996)
- Black & Wiliam (1998)
- Nyquist (2003)
- Brookhart (2004)
- Allal & Lopez (2005)
- Köller (2005)
- Brookhart (2007)
- Wiliam (2007)
- Hattie & Timperley (2007)
- Shute (2008)

## The effectiveness issue

## Understanding meta-analysis

31

- A technique for aggregating results from different studies by converting empirical results to a common measure (usually effect size)
- Standardized effect size is defined as:

$$\frac{experimental\ mean - control\ group\ mean}{population\ standard\ deviation}$$

- Problems with meta-analysis
  - The "file drawer" problem
  - Variation in population variability
  - Selection of studies
  - Sensitivity of outcome measures

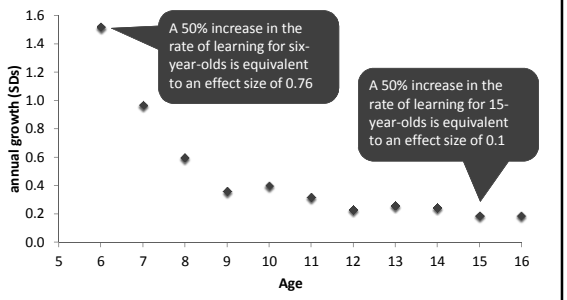
## The importance of statistical power

- The statistical power of an experiment is the probability that the experiment will yield an effect that is large enough to be statistically significant.
- In single-level designs, power depends on
  - significance level set
  - magnitude of effect
  - size of experiment
- The power of most social studies experiments is low
  - Psychology: 0.4 (Sedlmeier & Gigerenzer, 1989)
  - Neuroscience: 0.2 (Burton et al., 2013)
  - Education: 0.4
- Only lucky experiments get published…

## Variation in variability

## Annual growth in achievement, by age

34




Bloom, Hill, Black, and Lipsey (2008)

## Variation in variability

35

- Studies with younger children will produce larger effect size estimates
- Studies with restricted populations (e.g., children with special needs, gifted students) will produce larger effect size estimates

## Selection of studies

## Feedback in STEM subjects

37

- Review of 9000 papers on feedback in mathematics, science and technology
- Only 238 papers retained
  - Background papers 24
  - Descriptive papers 79
  - Qualitative papers 24
  - Quantitative papers 111
    - Mathematics 60
    - Science 35
    - Technology 16

Ruiz-Primo and Li (2013)

## Classification of feedback studies

38

1. Who provided the feedback (teacher, peer, self, or technology-based)?
2. How was the feedback delivered (individual, small group, or whole class)?
3. What was the role of the student in the feedback (provider or receiver)?
4. What was the focus of the feedback (e.g., product, process, self-regulation for cognitive feedback; or goal orientation, self-efficacy for affective feedback)
5. On what was the feedback based (student product or process)?
6. What type of feedback was provided (evaluative, descriptive, or holistic)?
7. How was feedback provided or presented (written, video, oral, or video)?
8. What was the referent of feedback (self, others, or mastery criteria)?
9. How, and how often was feedback given in the study (one time or multiple times; with or without pedagogical use)?

## Main findings

39

| Characteristic of studies included | Maths | Science |
|---|---|---|
| Feedback treatment is a single event lasting minutes | 85% | 72% |
| Reliability of outcome measures | 39% | 63% |
| Validity of outcome measures | 24% | 3% |
| Dealing only or mainly with declarative knowledge | 12% | 36% |
| Schematic knowledge (e.g., knowing why) | 9% | 0% |
| Multiple feedback events in a week | 14% | 17% |
| | | |

## Sensitivity to instruction

## Sensitivity of outcome measures

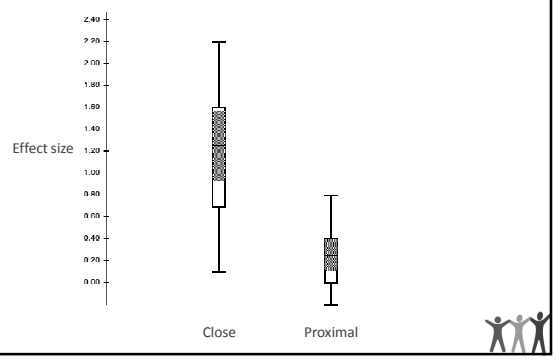
41

- Distance of assessment from the curriculum
  - Immediate
    - e.g., science journals, notebooks, and classroom tests
  - Close
    - e.g., where an immediate assessment asked about number of pendulum swings in 15 seconds, a close assessment asks about the time taken for 10 swings
  - Proximal
    - e.g., if an immediate assessment asked students to construct boats out of paper cups, the proximal assessment would ask for an explanation of what makes bottles float
  - Distal
    - e.g., where the assessment task is sampled from a different domain and where the problem, procedures, materials and measurement methods differed from those used in the original activities
  - Remote
    - standardized national achievement tests.

Ruiz-Primo, Shavelson, Hamilton, and Klein (2002)

## Impact of sensitivity to instruction

42



## Recent meta-analytic findings

| Content area | N | 95% confidence interval for effect size | | |
|---|---|---|---|---|
| | | Lower | Mean | Upper |
| Mathematics | 19 | 0.14 | 0.17 | 0.20 |
| English Language Arts | 4 | 0.30 | 0.32 | 0.34 |
| Science | 17 | 0.06 | 0.19 | 0.31 |
| Total | 40 | | | |

Mean effect size ≈ 0.20
A *big* effect size
Equivalent to a 50% to 70% increase in the rate of learning

Kingston and Nash (2011, 2015)

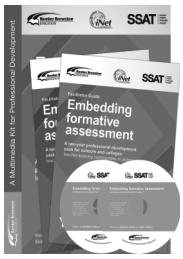
## Comments? Questions?

www.dylanwiliam.net

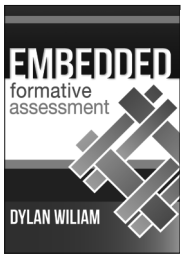
# RELATED RESOURCES

## Available from Hawker Brownlow Education

| Qty | Code | Title | Price |
|---|---|---|---|
| | SAT8899 | Embedding Formative Assessment Professional Development Pack | $544.50 |
| | SOT8112 | Embedded Formative Assessment | $35.95 |
| | GLA1284 | Inside the Black Box | $10.95 |
| | GLA1291 | Working Inside The Black Box | $10.95 |
| | GLA1345 | Inside the Black Box: Geography | $10.95 |
| | GLA1369 | Inside The Black Box of Assessment | $10.95 |
| | GLA1314 | Inside the Black Box: English | $10.95 |
| | GLA1376 | Inside the Black Box: Foreign Languages | $10.95 |
| | GLA1307 | Inside the Black Box: Primary Years | $10.95 |
| | GLA1338 | Inside the Black Box: Science | $10.95 |
| | GLA1383 | Inside the Black Box: Design and Digital Technologies | $10.95 |
| | GLA1321 | Inside the Black Box: Maths | $10.95 |
| | GLA1280 | Inside the Black Box Series Set of 11 | $110.00 |
| | | **Total (plus freight) $** | |

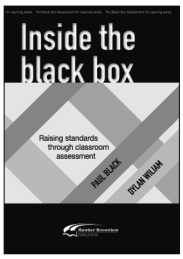


SAT8899

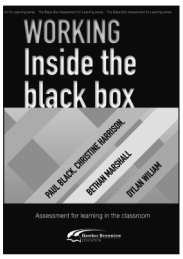


SOT8112

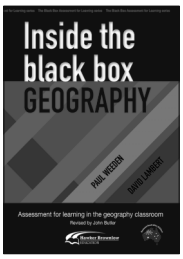
GLA1280



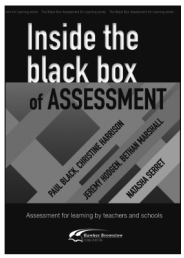
GLA1284



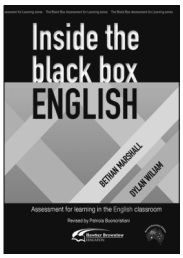
GLA1291



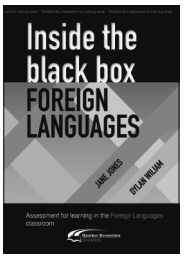
GLA1345



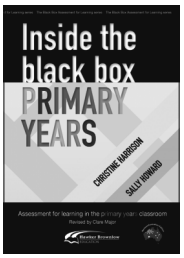
GLA1369



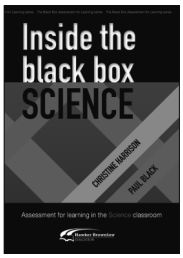
GLA1314



GLA1376

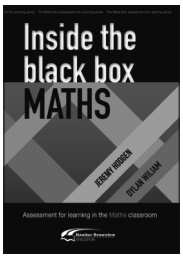


GLA1307



GLA1338

GLA1383



GLA1321

Attention ..... Order Number .....

Name of School.....

Address.....

..... State .....P/Code .....

Country .....

Email:.....

☐ Yes, I would like to receive emails from Hawker Brownlow Education about future workshops, conferences and the latest publications.

### Terms of Trade

- Prices are quoted in Australian dollars ($AUD) and include GST
- All prices are subject to change without notice.
- For New Zealand customers, at the time of invoice, we will convert the amount into New Zealand dollars ($NZD) so that you can pay by cheque or credit card in New Zealand dollars ($NZD).
- Full money-back guarantee.
- We do realise it is difficult to order sight unseen. To assist you in your selection, please visit our website <www.hbe.com.au>. Go to 'Browse Books' and most titles will give you the option to view the first few pages of the book. Click 'View Contents' on your selected book page.
- We will supply our books on approval, and if they do not suit your requirements we will accept undamaged returns for full credit or refund. Posters are for firm sale only and will not be sent on approval. Please be aware that delivery and return postage is the responsibility of the customer.
- Freight costs are determined at Australia Post rates, with a minimum delivery charge of $9.50 within Australia and $15.00 for New Zealand for each order.
- Please provide your street address for delivery purposes.

## To place an order, request a catalogue or find out more about our resources:

Call
1800 334 603
(03) 8558 2444

Fax
1800 150 445
(03) 8558 2400

Online
www.hbe.com.au

Mail
Hawker Brownlow Education
PO Box 580,
Moorabbin, VIC 3189

Do you want to know all about the latest professional development events in your area? Be the first to find out about new releases from world-renowned and local authors with the HBE e-newsletter! Upcoming titles will feature authentic assessment and digital media, along with a strong focus on success in mathematics and literacy. Sign up to our FREE e-newsletter at www.hbe.com.au.

**Online 'On Account' ordering now available!**
If you have a pre-existing account with Hawker Brownlow Education, you can now order online and pay using that account.