From ten classrooms to ten thousand: heuristics for scaling up formative assessment

Dylan Wiliam

Presentation at the annual meeting of the Association for Educational Assessment-Europe, November 2009: Malta

www.ioe.ac.uk

Overview

The evolving definition of formative assessment

Three heuristics for scaling up formative assessment

⌘ Decision-pull rather than data-push

⌘ Tight but loose

⌘ Content, then process

Putting all this into practice

Relevant studies

Fuchs & Fuchs (1986)

Natriello (1987)

Crooks (1988)

Bangert-Drowns, et al. (1991)

Kluger & DeNisi (1996)

Black & Wiliam (1998)

Nyquist (2003)

Dempster (1991, 1992)

Elshout-Mohr (1994)

Brookhart (2004)

Allal & Lopez (2005)

Köller (2005)

Brookhart (2007)

Wiliam (2007)

Hattie & Timperley (2007)

Shute (2008)

Definitions of formative assessment

We use the general term *assessment* to refer to all those activities undertaken by teachers—and by their students in assessing themselves—that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes *formative assessment* when the evidence is actually used to adapt the teaching to meet student needs" (Black & Wiliam, 1998 p. 140)

"the process used by teachers and students to recognise and respond to student learning in order to enhance that learning, during the learning" (Cowie & Bell, 1999 p. 32)

"assessment carried out during the instructional process for the purpose of improving teaching or learning" (Shepard et al., 2005 p. 275)

"Formative assessment refers to frequent, interactive assessments of students' progress and understanding to identify learning needs and adjust teaching appropriately" (Looney, 2005, p. 21)

"A formative assessment is a tool that teachers use to measure student grasp of specific topics and skills they are teaching. It's a 'midstream' tool to identify specific student misconceptions and mistakes while the material is being taught" (Kahl, 2005 p. 11)

"Assessment for Learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (Broadfoot et al., 2002 pp. 2-3)

Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence. An assessment activity can help learning if it provides information that teachers and their students can use as feedback in assessing themselves and one another and in modifying the teaching and learning activities in which they are engaged. Such assessment becomes "formative assessment" when the evidence is actually used to adapt the teaching work to meet learning needs. (Black et al., 2004 p. 10)

Formative assessment: a definition

"An assessment functions formatively to the extent that evidence about student achievement elicited by the assessment is interpreted and used to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions that would have been taken in the absence of that evidence.

Formative assessment therefore involves the creation of, and capitalization upon, moments of contingency (short, medium and long cycle) in instruction with a view to regulating learning (proactive, interactive, and retroactive)." (Wiliam, 2009)

The formative assessment hi-jack...

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: 5 seconds to 2 hours

⌘ Impact: classroom practice; student engagement

Unpacking assessment for learning

Key processes

⌘Establishing where the learners are in their learning

⌘Establishing where they are going

⌘Working out how to get there

Participants

⌘Teachers

⌘Peers

⌘Learners

Aspects of assessment for learning

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

Five "key strategies"…

Clarifying, understanding, and sharing learning intentions

⌘curriculum philosophy

Engineering effective classroom discussions, tasks and activities that elicit evidence of learning

⌘classroom discourse, interactive whole-class teaching

Providing feedback that moves learners forward

⌘ feedback

Activating students as learning resources for one another

⌘ collaborative learning, reciprocal teaching, peer-assessment

Activating students as owners of their own learning

⌘metacognition, motivation, interest, attribution, self-assessment

(Wiliam & Thompson, 2007)

…and one big idea

Use evidence about learning to adapt instruction to better meet learner needs

Tight but loose

www.ioe.ac.uk

Opposing factors in school reform

Need for flexibility to adapt to local constraints and affordances

⌘Implies there is appropriate flexibility built into the reform

Need to maintain fidelity to the theory of action of the reform, to minimise "lethal mutations"

⌘So you have to have a clearly articulated theory of action

Different innovations have different approaches to flexibility

⌘Some reforms are too loose (e.g., the 'Effective schools' movement)

⌘Others are too tight (e.g., Montessori Schools)

The "tight but loose" formulation

… combines an obsessive adherence to central design principles (the "tight" part) with accommodations to the needs, resources, constraints, and affordances that occur in any school or district (the "loose" part), but only where these do not conflict with the theory of action of the intervention.

A model for teacher learning

Content, then process

Content (what we want teachers to change)

⌘Evidence

⌘Ideas (strategies and techniques)

Process (how to go about change)

⌘Choice

⌘Flexibility

⌘Small steps

⌘Accountability

⌘Support

Choice

Belbin inventory (Management teams: why they succeed or fail)

⌘ Eight team roles (defined as "A tendency to behave, contribute and interrelate with others in a particular way.")

 ⊡ Company worker; Innovator; Shaper; Chairperson; Resource investigator; Monitor/evaluator; Completer/finisher; Team worker

⌘ Key ideas

 ⊡ Each role has strengths and allowable weaknesses

 ⊡ People rarely sustain "out of role" behavior, especially under stress

Each teacher's personal approach to teaching is similar

⌘ Some teachers' weaknesses require immediate attention

⌘ For most, however, students benefit more by developing teachers' strengths

# Flexibility

Distinction between strategies and techniques

⌘Strategies define the territory of formative assessment (no brainers)

⌘Teachers are responsible for choice of techniques

⌂Allows for customization/ caters for local context

⌂Creates ownership
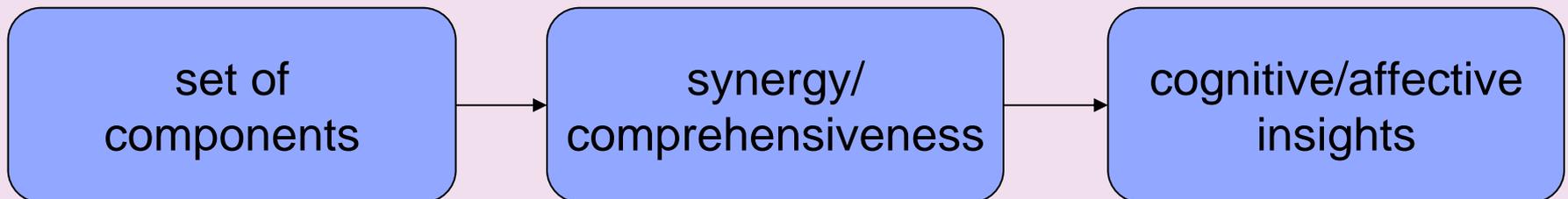
⌂Shares responsibility

Key requirements of techniques

⌘embodiment of deep cognitive/affective principles

⌘relevance

⌘feasibility

⌘acceptability

Design and intervention

Our design process

cognitive/affective insights	→	synergy/ comprehensiveness	→	set of components
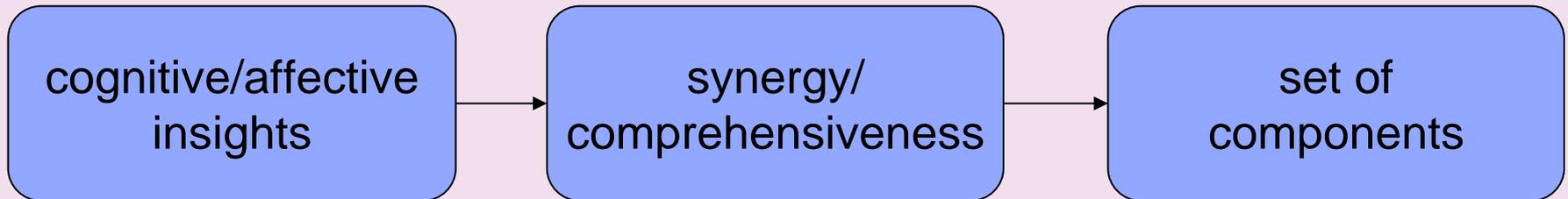
Teachers' implementation process

set of components	→	synergy/ comprehensiveness	→	cognitive/affective insights

Small steps

According to Berliner (1994), experts

⌘ excel mainly in their own domain.

⌘ often develop automaticity for the repetitive operations that are needed to accomplish their goals.

⌘ are more sensitive to the task demands and social situation when solving problems.

⌘ are more opportunistic and flexible in their teaching than novices.

⌘ represent problems in qualitatively different ways than novices.

⌘ have fast and accurate pattern recognition capabilities. Novices cannot always make sense of what they experience.

⌘ perceive meaningful patterns in the domain in which they are experienced.

⌘ begin to solve problems slower but bring richer and more personal sources of information to bear on the problem that they are trying to solve.

Example: CPR (Klein & Klein, 1981)

Six video extracts of a person delivering cardio-pulmonary resuscitation (CPR)

⌘5 of the video extracts are students

⌘1 of the video extracts is an expert

Videos shown to three groups: students, experts, instructors

Success rate in identifying the expert:

⌘Experts: 90%

⌘Students: 50%

⌘Instructors: 30%

Looking at the wrong knowledge…

The most powerful teacher knowledge is not explicit

⌘ That's why telling teachers what to do doesn't work

⌘ What we know is more than we can say

⌘ And that is why most professional development has been relatively ineffective

Improving practice involves changing habits, not adding knowledge

⌘ That's why it's hard

☐ And the hardest bit is not getting new ideas into people's heads

☐ It's getting the old one's out

⌘ That's why it takes time

But it doesn't happen naturally

⌘ If it did, the most experienced teachers would be the most productive, and that's not true (Hanushek, 2005)

Support

What is needed from teachers

⌘A commitment to:

⌃the continuous improvement of practice

⌃focus on those things that make a difference to student outcomes

What is needed from leaders

⌘A commitment to:

⌃creating expectations for the continuous improvement of practice

⌃ensuring that the the focus stays on those things that make a difference to student outcomes

⌃providing the time, space, dispensation and support for innovation

⌃supporting risk-taking

A case study in risk

Transposition of the great arteries (TGA)

⌘ A rare, but extremely serious, congenital condition in newborn babies (~25 per 100,000 live births) in which

⊟ the aorta emerges from the right ventricle and so receives oxygen-poor blood, which is carried back to the body without receiving more oxygen

⊟ the pulmonary artery emerges from the left ventricle and so receives the oxygen-rich blood, which is carried back to the lungs

⌘ Traditional treatment: the 'Senning' procedure which involves:

⊟ the creation of a 'tunnel' between the ventricles, and

⊟ the insertion of a 'baffle' to divert oxygen-rich blood from the left ventricle (where it shouldn't be) to the right ventricle (where it should)

⌘ Prognosis

⊟ Early death rate (first 30 days): 12%

⊟ Life expectancy: 46.6 years

The introduction of the 'switch' procedure

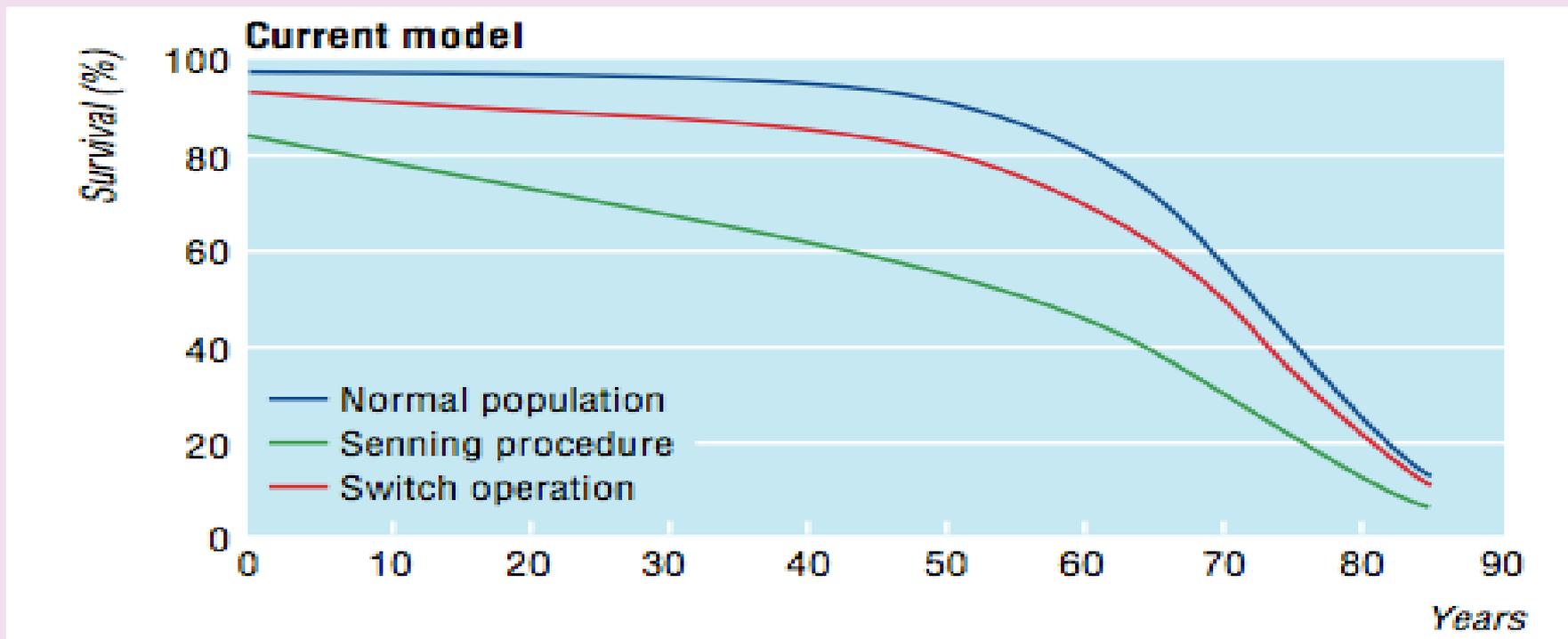Senning   Transitional   Switch

Transposition of great arteries
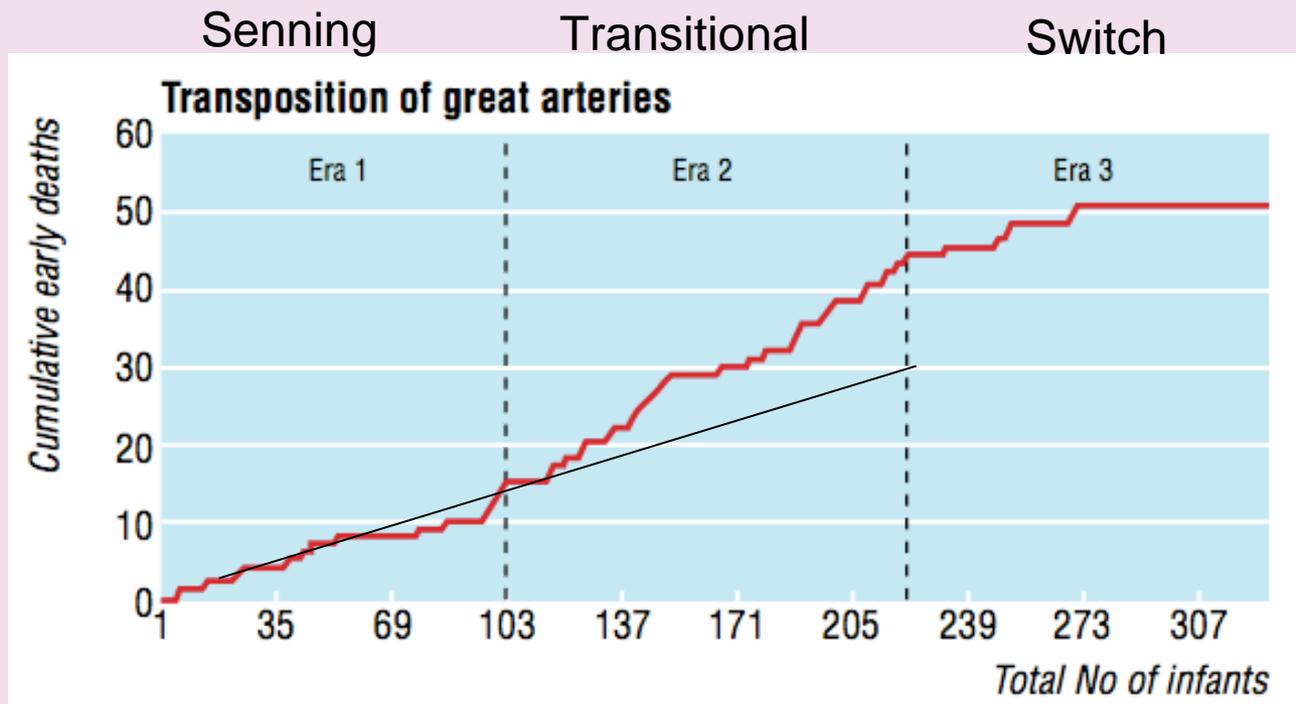
Early death rate
Senning   12%
Transitional   25%   Bull, et al (2000). *BMJ,* **320**, 1168-1173.

Impact on life expectancy

Current model

Survival (%) vs Years

- Normal population
- Senning procedure
- Switch operation

Life expectancy:
Senning: 46.6 years
Switch: 62.6 years

Making a commitment…

Action planning

⌘ Forces teachers to make their ideas concrete and creates a record

⌘ Makes the teacher accountable for doing what they promised

⌘ Requires each teacher to focus on a small number of changes

⌘ Requires the teacher to identify what they will give up or reduce

A good action plan

⌘ Does not try to change everything at once

⌘ Spells out specific changes in teaching practice

⌘ Relates to the five "key strategies" of AfL

⌘ Is achievable within a reasonable period of time

⌘ Identifies something that the teacher will no longer do or will do less of

…and being held to it

I think specifically what was helpful was the ridiculous NCR forms. I thought that was the dumbest thing, but I'm sitting with my friends and on the NCR form I write down what I am going to do next month.

Well, it turns out to be a sort of "I'm telling my friends I'm going to do this" and I really actually did it and it was because of that. It was because I wrote it down

I was surprised at how strong an incentive that was to do actually do something different … that idea of writing down what you are going to do and then because when they come by the next month you better take out that piece of paper and say"Did I do that?" … just the idea of sitting in a group, working out something, and making a commitment… I was impressed about how that actually made me do stuff. (Tim, Spruce Central High School)

Support

Teacher learning is just like any other learning in a highly complex area

⌘In the same way that teachers cannot do the learning for their learners

⌘Leaders cannot do the learning for their teachers

Two extreme responses

⌘"It's hopeless"

⌘Let a thousand flowers bloom..

Neither will work

⌘What leaders can do is engineer effective learning environments for teachers

⌘'Servant' leadership

Progress of TLCs in Cannington

	Maths	Science	MFL
Ash	1 —	1 —	0 —
Cedar	5 ▮	1 ▮	3 ▮▮
Hawthorne	4 ▮▮	10 ▮▮	5 ▮▮▮▮
Hazel	7 —	12 —	2 —
Larch	1 ▮▮▮▮▮	0 ▮	0 ▮
Mallow	6 ▮▮▮	7 ▮	3 ▮▮
Poplar	11 ▮	3 ▮▮▮	1 ▮▮▮
Spruce	7 ▮▮▮▮	8 ▮▮▮	5 ▮▮▮
Willow	2 ▮	5 ▮	2 ▮▮▮▮
Totals	44	47	21

Black nos. show teachers attending launch event; blue bars show progress of TLC

Pareto analysis

Vilfredo Pareto (1848-1923)

⌘ Economist, philosopher, etc., associated with the 80:20 rule

Pareto improvement

⌘ A change that can make at least one person (e.g., a student) better off without making anyone else (e.g., a teacher) worse off.

Pareto efficiency/Pareto optimality

⌘ An allocation (e.g., of resources) is Pareto efficient or Pareto optimal when there are no more Pareto improvements

Schools are rarely Pareto optimal

Examples of Pareto improvements

⌘ Less time on marking to spend more time on planning questions to use in lessons

⌘ Increased use of peer assessment

Obstacles to Pareto improvements

⌘ The political economy of reform

⌘ In professional settings, it is relatively easy to stop people doing bad things

⌘ It is incredibly hard to stop people doing valuable things in order to give them time to do *even more* valuable things

◹ e.g., "Are you saying what I am doing is no good?"

The story so far…

1993-1998

⌘ Review of research on formative assessment

1998-2003

⌘ Face-to-face implementations with small groups of teachers

⌘ Effect sizes ~0.3 standard deviations (equivalent to > 50% increase in rate of learning)

2003-2008

⌘ Attempts to produce faithful implementations at scale

2008-2013

⌘ Creating the conditions for implementations at scale

⌘ Making formative assessment the *only* number one priority…