Assessment, learning and technology: prospects at the periphery of control

Keynote address to ALT-C
September 2007; Nottingham, UK

Dylan Wiliam
Institute of Education,
University of London

www.dylanwiliam.net
Overview of presentation

Theoretical precepts
• About learning
• About teaching
  – Pedagogies of engagement
  – Pedagogies of contingency

The role of technology
• Supporting, rather than replacing, teachers
• Classroom aggregation technologies
Raising achievement matters

For individuals
• Increased lifetime salary
• Improved health
• Longer life

For society
• Lower criminal justice costs
• Lower health-care costs
• Increased economic growth
Where’s the solution?

Structure
• Smaller high schools
• Larger high schools
• K-8 schools

Alignment
• Curriculum reform
• Textbook replacement

Governance
• Charter schools
• Vouchers

Technology
• Computers
• Interactive white-boards
School effectiveness

Three generations of school effectiveness research

• Raw results approaches
  – Different schools get different results
  – Conclusion: Schools make a difference

• Demographic-based approaches
  – Demographic factors account for most of the variation
  – Conclusion: Schools don’t make a difference

• Value-added approaches
  – School-level differences in value-added are relatively small
  – Classroom-level differences in value-added are large
  – Conclusion: An effective school is a school full of effective classrooms
It’s the classroom

Variability at the classroom level is up to 4 times that at school level
It’s not class size
It’s not the between-class grouping strategy
It’s not the within-class grouping strategy
It’s the teacher
Teacher quality

A labour force issue with 2 solutions

• Replace existing teachers with better ones?
  – No evidence that more pay brings in better teachers
  – No evidence that there are better teachers out there deterred by burdensome certification requirements

• Improve the effectiveness of existing teachers
  – The “love the one you’re with” strategy
  – It can be done
  – We know how to do it, but at scale? Quickly? Sustainably?
How do we improve teaching?
# Quality control and quality assurance

## Quality control
- Bolt-on
- Determines need for re-processing
- Quality is “inspected in”
- Bad

## Quality assurance
- Built-in
- Obviates the need for re-processing
- Quality is “designed in”
- Good
Except that…

For some process quality assurance is more efficient than quality control
• e.g., automobile manufacture

For some process quality control is more efficient than quality assurance
• e.g., silicon chip manufacture

Crucial trade-offs: testability vs complexity vs predictability

Where does learning fit?
What gets learnt?

(Denvir & Brown, 1986)
What gets learned (2)?

Which fraction is the smallest?

a) \( \frac{1}{6} \), b) \( \frac{2}{3} \), c) \( \frac{1}{3} \), d) \( \frac{1}{2} \).

Success rate 88%

Which fraction is the largest?

a) \( \frac{4}{5} \), b) \( \frac{3}{4} \), c) \( \frac{5}{8} \), d) \( \frac{7}{10} \).

Success rate 46%; 39% chose (b)

(Vinner, PME conference, Lahti, Finland, 1997)
What gets learned (3)?

860 + 570 =

(Leverhulme Numeracy Research Programme)
What gets learned (4)?

Strategies & errors in secondary mathematics (Booth; 1984; Hart, 1984)

- One-third knew the content at the beginning
- One-third didn’t know the content at the end
- One-third learnt the content
- But, half of these had forgotten the content six weeks later
- However, some did better on the delayed post-test than on the immediate post-test
Key insights from C20th psychology

1. What gets learned as a result of a particular sequence of instructional activities is impossible to predict, but
2. Student errors are not random

Conclusion: teaching is interesting because learners are so different, but only possible because they are so similar

Learning is a liminal process, at the boundary between control and chaos
Learning power environments

Key concept:
• Teachers do not create learning
• Learners create learning

Teaching as engineering learning environments

Key features:
• Create student engagement (pedagogies of engagement)
• Well-regulated (pedagogies of contingency)
Why pedagogies of engagement?

Intelligence is partly inherited
• So what?

Intelligence is partly environmental
• Environment creates intelligence
• Intelligence creates environment

Learning environments
• High cognitive demand
• Inclusive
• Obligatory
Motivation: cause or effect?

(Csikszentmihalyi, 1990)
Why pedagogies of contingency?

For evaluating institutions
For describing individuals
For supporting learning

• Monitoring learning
  – Whether learning is taking place
• Diagnosing (informing) learning
  – What is not being learnt
• Forming learning
  – What to do about it
Effects of formative assessment

Several major reviews of the research

• Natriello (1987)
• Crooks (1988)
• Kluger & DeNisi (1996)
• Black & Wiliam (1998)
• Nyquist (2003)

All find consistent, substantial effects
Kinds of feedback (Nyquist, 2003)

Weaker feedback only
• Knowledge of results (KoR)

Feedback only
• KoR + clear goals or knowledge of correct results (KCR)

Weak formative assessment
• KCR+ explanation (KCR+e)

Moderate formative assessment
• (KCR+e) + specific actions for gap reduction

Strong formative assessment
• (KCR+e) + activity
## Effect of formative assessment (HE)

<table>
<thead>
<tr>
<th>Type of Assessment</th>
<th>N</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weaker feedback only</td>
<td>31</td>
<td>0.14</td>
</tr>
<tr>
<td>Feedback only</td>
<td>48</td>
<td>0.36</td>
</tr>
<tr>
<td>Weaker formative assessment</td>
<td>49</td>
<td>0.26</td>
</tr>
<tr>
<td>Moderate formative assessment</td>
<td>41</td>
<td>0.39</td>
</tr>
<tr>
<td>Strong formative assessment</td>
<td>16</td>
<td>0.56</td>
</tr>
</tbody>
</table>

(Nyquist, 2003; revised values)
## Cost/effect comparisons

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Extra learning</th>
<th>Cost/yr/classroom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class-size reduction (by 30%)</td>
<td>20%</td>
<td>£20k</td>
</tr>
<tr>
<td>Increase teacher content knowledge by 1 sd</td>
<td>5%</td>
<td>?</td>
</tr>
<tr>
<td>Formative assessment/Assessment for learning</td>
<td>75%</td>
<td>£2k</td>
</tr>
</tbody>
</table>
Three generations of pedagogy

First generation
- Traditional pedagogy
- Negligible contingency

Second generation
- All student response systems
- Contingency dependent entirely on teacher skill

Third generation
- Automated aggregation technologies
- Contingency supported by technology
Four-process architecture

Task selection
Task presentation
Evidence identification
Evidence accumulation

Almond, Steinberg and Mislevy (2002)
Task selection/
Task presentation
Questioning in math: discussion

Look at the following sequence:
3, 7, 11, 15, 19, ....

Which is the best rule to describe the sequence?

A. n + 4
B. 3 + n
C. 4n - 1
D. 4n + 3
In which of these right triangles is $a^2 + b^2 = c^2$?
Questioning in science: discussion

Ice-cubes are added to a glass of water. What happens to the level of the water as the ice-cubes melt?

A. The level of the water drops
B. The level of the water stays the same
C. The level of the water increases
D. You need more information to be sure
The ball sitting on the table is not moving. It is not moving because:

A. no forces are pushing or pulling on the ball.
B. gravity is pulling down, but the table is in the way.
C. the table pushes up with the same force that gravity pulls down
D. gravity is holding it onto the table.
E. there is a force inside the ball keeping it from rolling off the table

(Wilson & Draney, 2004)
Save the ozone layer

What can we do to preserve the ozone layer?

A. Reduce the amount of carbon dioxide produced by cars and factories
B. Reduce the greenhouse effect
C. Stop cutting down the rainforests
D. Limit the numbers of cars that can be used when the level of ozone is high
E. Properly dispose of air-conditioners and fridges
Questioning in English: discussion

Macbeth: mad or bad?
Questioning in English: diagnosis

Where is the verb in this sentence?

The dog ran across the road

A B C D
Questioning in English: diagnosis

Which of these is the best thesis statement?

A. The typical TV show has 9 violent incidents
B. There is a lot of violence on TV
C. The amount of violence on TV should be reduced
D. Some programs are more violent than others
E. Violence is included in programs to boost ratings
F. Violence on TV is interesting
G. I don’t like the violence on TV
H. The essay I am going to write is about violence on TV
Hinge Questions

A hinge question is based on the important concept in a lesson that is critical for students to understand before you move on in the lesson.

The question should fall about midway during the lesson. Every student must respond to the question within two minutes. You must be able to collect and interpret the responses from all students in 30 seconds.
Figurative language

A. Alliteration
B. Hyperbole
C. Irony
D. Metaphor
E. Onomatopoeia
F. Personification
G. Simile
H. None of the above

1. He was a bull in a china shop.
2. May I have a drop of water?
3. This backpack weighs a ton.
4. The sweetly smiling sunshine...
5. He honked his horn at the cyclist.
6. I’ve told you a million times already.
7. The Redcoats are coming!
8. “They in the sea being burnt, they in the burnt ship drown’d.”
9. He was as tall as a house.
Evidence identification

Single student response systems

All-student response systems

- Flash-cards/dry erase boards
- Classroom ‘clickers’
- Traditional keyboards (wired/wireless)
- Anoto pens
Anoto pen

Wireless pen
Special coated paper
Pen ‘knows where it is’
Palm with wireless keyboard

- Text-based input
- Limited task-presentation capability
- Portable
Classroom ‘clickers’
A big meteor from space hit the Earth. It threw a great cloud of dust that covered the sun worldwide. Plants died and dinosaurs starved. Earth’s climate or the weather changed which affected what they ate. Their brains were so small so they could not adapt and they died.

| Angola, Oliver | 2 | postal service |
| Dan, Walkerman | 0 | 1 | postal service |
| Davis, Nicole | 0 | 2 | postal service |
| Erdosy, Judi | 0 | 1 | USPS |
| Maggie, Fincher | 0 | 3 | distribution |
| Mariano, Carolyn | 0 | postal service |
| Nancy, Fenyok | 0 | postal service |
| Ria, Chehade | 0 | 1 | The New York Times |
| Val, Zhute | 0 | 1 | “which” refers back to “postal service” (btw-this is a very poorly phrased sentence...) |
Evidence identification

Automated essay scoring (e-rater)
Paraphrase analysers (c-rater)
Graphical and equation analysers (m-rater)
Evidence identification technologies

- clickers
- m-rater
- c-rater
- e-rater
- ABCD cards
- dry-erase boards
- Discourse®

- automated
- unstructured
- aggregated
- structured
- teacher-mediated
- latent semantic analysis

Evidence structure
Evidence accumulation

Unidimensional student models
• Useful for summative purposes
• Almost useless for formative purposes

Multidimensional student models

Evidence-centred design
• Bayesian inference networks
  – Proficiency model
  – Task model
  – Evidence model
  – Student model
Evidence utilization

Whole-class

Sub-groups
  • Homogenous
  • Heterogenous

Individualization
Summary

Raising achievement is important
To do so, we have to change what happens in classrooms
We have to work with, rather than replace teachers
Specifically, it is more important to improve pedagogy than subject matter knowledge
• Pedagogies of engagement
• Pedagogies of contingency
  – Single-student response systems
  – All-student response systems
  – Classroom aggregation technologies