EVS: linked questions; meta-messages

(MCQ = Multiple Choice Questions
EVS = Electronic Voting Systems)

Steve Draper, Glasgow University

www.psy.gla.ac.uk/~steve/talks/evs9.html

1

Part A: Prelude

2

What do we want students to learn?

What is it that we most want our students to learn / acquire?

In general, in any discipline, they (we) implicitly want students to learn to think like an X-icist: a physicist, a historian, a sociologist, a psychologist.

But what does this really mean?
To know facts, to solve equations, to .... what exactly?

I will end this talk by arguing that "this" may be a type or style of conversation or discussion.

Not the questions but ...

The focus here is that what matters:

Is not (just) the design of individual EVS questions

But what the presenter does with them that matters.

And that very often means:

How multiple questions are related to each other.

Many good EVS techniques use linked questions

- Class tests: students work through a set of questions; then bulk entry and feedback.
- Stepwise problem-solving (one vote for each step).
- Diagnostic tree of questions (e.g. in a revision session)
- Re-use of same data bank for questions in different directions (e.g. "which of these is the symbol for iron?: Au, Fe .." vs. "Which of these is the name for the element Hg?: Iron, silver, mercury?")
- Bowskill's induction sessions: collect student "concerns", then rate each one. (First they create the questions; then everyone answers.)
- Mark Russell's diagnostic triplets
- Mazur's PI (peer instruction): vote, discuss, re-vote.

Part B1:

Several linked questions per concept

A learner can get an MCQ right when they don't really understand the concept it is trying to test

6

Linked questions per concept

A student can get a 4-option MCQ right by chance 25% of the time.
And even with no guessing, getting one example of a concept right often does not predict that that student can apply it in a different question (transfer).

Mark Russell: superior tactic of linked questions for diagnosis.
"Using an electronic voting system to enhance learning and teaching" Mark Russell (2008) Engineering Education vol.3 no.4 pp.58-65

N ≈ 71 students
Three questions testing same underlying concept of pressure gradient (how pressure varies with depth in water).

7

Mark Russell (1)

Qu. 1 The pressure gradient (dp/dz) in a fluid ...

a) is always zero
b) is always negative
c) is always positive
d) can be either depending on the fluid
e) is equal to the gauge pressure

8

See figs.3-9 in:

Mark Russell (2008) "Using an electronic voting system to enhance learning and teaching" Engineering Education vol.3 no.4 pp.58-65

Mark Russell diagrams

9

The Mark Russell case as numbers

	Q1	Q3	Q2	1&2	1&3	2&3		All 3
% correct	83%	63%	58%	50%	51%	?		36%

The argument is that the conjunction of questions is a more powerful test of understanding than any one by itself.

Part B2:

Several linked questions per concept

Peer discussion; and showing the learning transfers to an isomorphic question on the concept

11

Smith et al. 2009 paper in Science

- IE / Mazur type method, but in level 1 Genetics, not physics
- Re-test was not only the identical question, but also another similar (isomorphic) one.
- Even when no-one knew the right answer, many students learned from the peer discussion (for 15 of 16 topics)
- Biggest improvement on the more difficult questions
- Delayed benefit in the sense that some got the isomorphic one right even if persisting in the wrong answer for the repeated question.

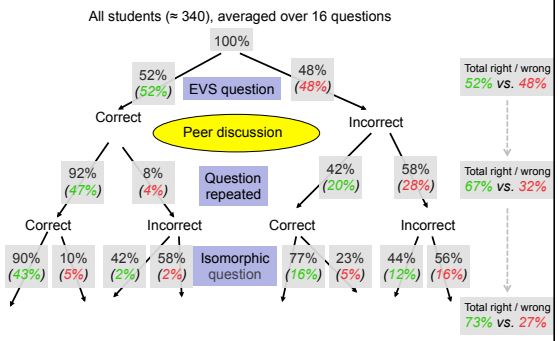
Went from 52% correct to 72.52% correct averaged over 16 qus.
(7.4% got worse; 28% better)

Big Result!

This is pretty good evidence that it isn't numerical scores but actual understanding (deep learning) that goes up, as evidenced by transfer to an isomorphic question.

And it is consistent the the "catalytic" explanation of the learning mechanism I discuss in a minute.

Redrawn from: **Smith et al. 2009 paper in Science**



Not the philosopher's stone, ...

So the Mazur recipe doesn't always work, in the sense that every learner improves their understanding after every discussion on every question.
Some learners seem to get worse.
Some questions seem to make a majority get worse.

But the overall pattern is strongly: a noisy random walk tending towards greater understanding in every class.

Here's some data from the first year a lecturer tried it, in a new subject with questions she wrote herself. (Her following year's data were better.)

15

Summary stats over the 58 questions on one course, each repeated after peer discussion	
88%	Percent of questions with some net gain
15%	Rise in student scores (average size of net gain, pre to post, as % of class)
32%	Average normalised gain (pre to post as % of those who got it wrong and could be improved) *[2/3rds of students fail to improve]*
93%	Percent of qus. where right answer wins in the end
16%	Percent of qus. where right answer comes from behind to win
57%	Percent of qus. where more than one option attracts and gains votes after discussion
57%	Average (mean) proportion of changed votes that change to the right answer. *[I.e. nearly half the changes aren't immediately successful. Lots of churn.]*

Part C: Interlude: Catalytic assessment

"Catalytic assessment" is a catch-phrase for questions that may look like tests, but whose important mathemagenic (learning-generating) effect is hidden in the learner.

17

Catalytic assessment

"Catalytic assessment" is a catch phrase for questions like Mazur's brain teasers that may look like tests, but whose important mathemagenic (learning generating) effect is hidden in the learner.
They are simple factual questions on the surface, but they make the learners think; and the result is far deeper understanding.

For the full argument see my paper on this:
"Catalytic assessment: understanding how MCQs and EVS can foster deep learning" British Journal of Educational Technology vol.40 no.2 pp.285-293

[This ref. is on the handout sheet]

18

Put simply

It's about getting learners to:
a) Consider whether they really understand this stuff (confidence self-assessment);
b) Generate reasons to help decide ("self-explanations")

Most people just don't think, so they don't learn very well: certainly, they don't work on their understanding, only on performing adequately. Get them to think, and the learning improves in quality and quantity.

19

Part D: Critiques of Powerpoint

These can be reviewed for clues about what can be good or bad qualities of lectures

Critiques of ppt

Turkle: school kids applaud each others' presentations on the basis of visual effects.
This poisons education by promoting an attitude that teaching is just clear summarisation, even of poetry.

Tufte: Ppt caused the Columbia disaster for just this reason: promoted simple summaries, not arguments or justifications.

Andeweg [CAL 07] experiment: people preferred pictures, but learned more from text slides. [A picture is worth 1/1000th of a word]

Tufte: Ppt is bad because it uses the visual channel badly. Talks become bad because they get designed around a bad use of this channel. It is not entertainment that most talks lack, but any real content.

Turkle: "Ppt encourages presentation not conversation".

Part E: The WordWall critique of ppt lectures

22

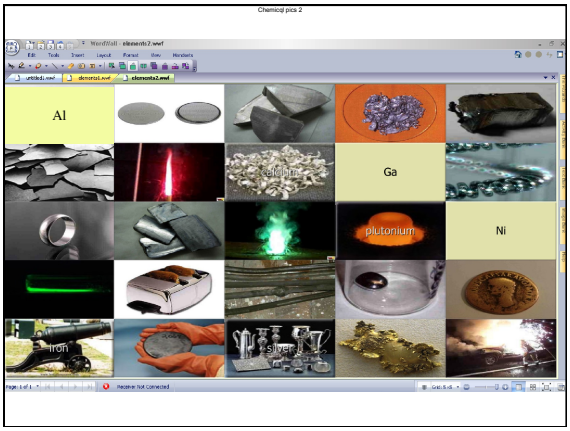
Challenge: Slideshows

When the founder of WordWall, a school teacher who left to create EVS equipment to meet the needs he'd felt as a teacher, was talking to us about possible entry into the HE market, we told him he'd have to make it compatible with Ppt for all those legacy sets of slides people had.

"But why would you want to give a slideshow instead of teaching?" [Ben Watson]

His idea of teaching is to create impromptu learning activities on the fly.

Palette vs. slideshow style of support
Does this imply we don't want a slideshow support like ppt but a palette-like toolbox for creating activities on the spot?



Slideshow
Narrative
Transmission
Telling
HE

25

Slideshow
Narrative
Transmission
Telling
HE

Palette
LBE
Constructivism
Teaching
Schools

26

Reprise

So at this stage of the argument, there are:

1. Bad instructivist slideshows – monologues
2. OR good, interactive EVS sessions which are interactive and engaging in these ways:
   a) Learners are active externally
   b) Teachers are responsive: vary what they do in response to the audience

Any questions so far?

28

Part F:
Norman Gray, Tufte, dense multi-modal presentations

29

Dense multi-modal presentations

The WordWall critique, Mazur's PI etc. can be read as "engaging" students on the spot; and providing responsive (contingent) teaching that reacts to the audience by on the spot activities.

A quite different phenomenon (but actually in line with Tufte's critique) is that the goal is exposition of complex content; of demanding that the audience work (be internally interactive), not that the presenter "engages" them externally.

E.g.1: Norman Gray's GR class

Norman teaches a final year option on General Relativity.

He tells me that his approach is to have dense printed "lecture notes" handed out, which have the details and equations exactly; but his oral performance (now recorded) has a casual style, and is a commentary on how students might approach the seamless, quasi-impenetrable, interlocking set of concepts and formalisms.

Thus his oral and written modes are not duplicates, nor entertainment, but a complementary 2-channel presentation.

E.g.2: Kant's "Critique of pure reason"

Norman's wife Susan Stuart is a career-long fan of Kant.

Kant's "critique" has a seamless mesh of interlocking arguments; and she is engaged in developing a "road map" widget to help philosophy students navigate it.

Again, the linear (or even tree) structure of narrative is a poor fit for expounding a rounded "system", which has no natural beginning or end.

(Maxwell's equations might be another example).

E.g.3: History lectures

At least according to my understanding of an informant, Vicky Gunn, history lectures traditionally have relied on a multi-modal (double-barrelled) communication.

Such lectures are put on, not as factual briefings, but as performances of what a respectable, but highly questionable, "performance" or argument in history can be. They make an argument the presenter may not believe, but which illustrate the way arguments are made, the ways theories are used as lenses (not axioms).

Originally, tutorial discussions would convey this attitude, this genre; and lectures were to be interpreted (only) in the light of that.

Conclusion from these cases

Tufte, of course, feels that the heart of a good presentation is some brilliantly crafted visual presentation of data. Full of things for the audience to think about, and meriting a lot of (unwritten) commentary by the presenter.

[A Tufte graphic merits a 1,000 words of commentary — but is not an alternative.]

Again, the twin barrels of visual and spoken complementing each other.

So:

There are thus alternative latent visions of what presentations, especially for learning, are about:

1. Shallow Powerpoint: "making a point not making an argument"
2. Contingent teaching: responding to the audience with on-the-fly activities
3. Multi-modal dense presentations: stretch the audience, not the presenter; Double-barrelled, linked documents.

But another, fourth, line of argument is possible ....

Any questions so far?

36

Part G:
The neo-Vygotskian argument

Meta-messages

37

The abstract idea

The wave-particle duality of Developmental Psychology is Piaget-Vygotsky. Recklessly summarising what I haven't read carefully, and boldly generalising from infancy to undergraduates, we have:

Piaget: the child is a little scientist, and constructs abstract theories of the world fromexperiments (playing with stuff). Other people are not really different from experiments: they may disturb the child's ideas, but it is the child who constructs the modifications. Congenial to scientists, and to the "catalytic" idea.

Vygotsky: All important methods of thought have their origins in conversations with adults. At first the child doesn't understand what they are doing, but are scaffolded by the adult. (Recite after me: eeny, meeny, miny, ...) They come to do more and more until they are equal partners in the interaction. Finally, they internalise it and what was interactive is now thought

Paul Black's argument

Attributed to a school teacher: who says he spends large amounts of science lesson class time in class discussions of what could explain various facts or phenomena, with no reference to the textbook or established knowledge.
"I can't assume the children have ever heard an argument based on who has the best reasons (as opposed to violence, shouting, appeals to authority, etc.) If they've never experienced this style of discussion, what possible sense could science make to them?"

I find this argument overwhelming.
It is, in essence, neo-Vygotskian.

What style of discussion is implied by our lessons?

Pub quizzes, didactic ppt: you just have to recall or guess; generative reasoning doesn't help.

Mazur's PI: Discuss reasons, not guess answers or assert opinions. That's how physics is done: reasoning it out.

Bowskill's induction recipe: It's OK to discuss your problems and worries as a student, with peers, and with staff.

We might call these "meta-messages": what sessions imply and enact (but don't state) about the kind of thinking, the kind of conversation, appropriate and required in this context.
For education, the meta-message could be more important than the "content".

(cont.)

This notion may also explain some of the details in what is going on in a PI/Mazur session.

When I do Reciprocal Peer Critiquing in my own tutorials, I get students to read each other's essays and feed back comments. The first time they do this, I get them to deliver the comments face to face with me present. This seems to establish the right tone: after that, they don't need me there. This is in contrast to other attempts (e.g. Aropa) where you hear many reports of rude and unhelpful comments.

In PI, Mazur and his GTAs patrol the room monitoring and steering the discussion. And may presenters seem to hide the vote results to stop students going with the majority. But the best brain teaser questions also significantly shift the discussion from eliminating distractors to reasoning about the theories.

(cont.)

Quintin and colleagues, however, seem to have more trouble with PI in computing science. They have the student "engagement"; but most of the conversations are not about compSci reasoning.

It's possible that common experience of IT leaves the students with the expectation that in computing, answers are arbitrary conventions and tricks, not things you could possibly reason out.

[Cf. Hall!]

Part H:
Conclusion, reprise on CDCs

(Core Disciplinary Criteria)

43

Core Disciplinary Criteria

Given a discipline, what do they/we want students to learn?
What is important for turning students into thinkers like us?

1. Facts and concepts

2. Core, complex procedures e.g. essay writing in sociology; equation manipulation in physics. We drill students in almost every exam or coursework on these; and take it as a measure of their worth in the subject

3. BUT: perhaps something else as well. Hestenes, with his FCI, created demonstrations which physics teachers could apply to their own classes, that showed devastating lack of ability at qualitative reasoning. One way of looking at this, is that this exposed a kind of thinking which physics teachers absolutely required, but hadn't realised they did require but needed to teach.

   It could be characterised as a kind of conversation; the quick arguments you'd give to yourself, or a colleague about a situation or problem. A way of doing thinking about physics.

A place to stop

For the slides, handout etc. see:

http://www.psy.gla.ac.uk/~steve/talks/evs9.html

45