Formative MCQs, Peer Interaction, and Deep Learning

Steve Draper,

School of Psychology,

Glasgow University

Bath Economics 9 May 2017

For the slides, handout etc. see:
http://www.psy.gla.ac.uk/~steve/talks/bath.html

1

Part 1:

Hunt; Gardner-Medwin

Just asking yourself whether you know something, prompts checking and deeper learning.

2

Asking about confidence (Hunt, 1982)

Hunt (1982) (in an artificial experiment) showed that participants who first chose an answer and then had to indicate a confidence level learned about 20% faster than those who just chose an answer.

This uses a setup where each learns in a cycle of trying to learn a list of pairs then doing the test; repeating till they reach 80% score.

Gardner-Medwin's CBM (confidence based marking) is a direct application of this.

➔ Just asking yourself whether you know something, prompts checking and deeper learning.

Pillar 1: Being aware of a hole in our knowledge prompts us to learn

Hunt's experiment seems to imply that:

- We can judge how well we know something
- We normally don't do this
- But we generally remedy it when we can (after becoming aware of a deficit).
- *(But in Howe's experiments this happened NOT at the time BUT some time between then and a 4 week follow-up.)*
- This underlies a number of different pedagogic designs, as we shall see.

Part 2:

Formative Tests:
their importance for causing learning

Mastery learning: Marks and remediation

5

The argument

Mastery Learning (ML) rests on the view that if the only tests a learner gets show differences between learners (but without comparing different teaching methods, learning actions, time taken, ….) then everyone tends to interpret them as about learner abilities.

What is poisonous about standard school and university teaching is to vary only the learner: so we are almost forced to interpret marks as about their ability. To learn effectively, instead, they need to monitor their learning after a first pass, and correct it: a totally different use of tests, with a different stance on capabilities.

6

Carroll and Mastery Learning

In 1963-1980 the work by J.B.Carroll and then Bloom showed the irrationality (and damaging nature) of the standard attitude that school tests measure ability.

If you assume the learning and teaching must be constant, then the spread of test scores looks like a measure of learner ability.

But (they showed) if you vary the time and/or teaching method, then the spread largely disappears: so the former spread can't be a measure of learner limitations.

ML set out to give every learner the experience, not of praise, but of objective success.

7

Mastery Learning: the method

Each subject is taught as usual (e.g. 1 hour per day) Mon-Thur.

On Friday, the session starts with a test; and the marks per question are fed back to each learner.

Learners then pick the parts they got wrong;
usually group with 2-3 other learners with the same problem;
and the groups then self-remediate.

Then everyone re-takes the test and receives their new marks.

8

Typical test scores

Summative achievement scores

9

Mastery learning scores

Summative achievement scores Mastery learning scores

10

ML: the summary

Tests, superficially seen as objective, are routinely mis-interpreted as a summative measure of learner ability.

Carroll and Bloom argued forcefully that the same test may be used formatively: as a guide for each individual to correct their learning, bringing it up the level of the best.

They demonstrated that this change in practice moved scores up by an average of one standard deviation.

Mastery Learning = Marks then remediation:
the formative use of tests as the central method for prompting learning.

Part 3:

Classic test of deep learning in Economics

Asking simple questions about everyday things frequently shows how students have made absolutely no change to their everyday understanding as the result of university teaching.

This demonstrates the lack of deep learning: how much learning is learning to repeat new things, while never thinking about whether it has any relationship to what you already "know".

12

A classic test of deep learning

Students leaving an introductory course on Economics were asked a simple open-ended question "Why does a loaf of bread cost one pound?"
On analysing the content of their answers, most were concerned to explain it in terms of the cost of wheat, the wages of the millers and bakers.
None related it to the law of supply and demand and how the price of bread will change in a famine.

[A paraphrase of a study reported by Dahlgren & Marton 1978]

The lack of deep learning -- Of student understanding

Asking simple questions about everyday things frequently shows how students have made absolutely no change to their everyday understanding as the result of university teaching.

This demonstrates the lack of deep learning: how much learning is learning to repeat new things, while never thinking about whether it has any relationship to what you already "know".

The widespread lack of thinking about what you have learned; what relationship it has to other knowledge you already have; whether you have understood it.

What I mean by "deep learning"

Shallow learning (no depth) is learning for a single foreseen use (test): most extremely, reproducing what was said with no processing.

Deep learning is connecting the new material to other things. The example above is connecting it to how the new idea relates to the learner's existing everyday experience. This is often what seems most lacking in HE learning.

However for me, depth is about creating more types of link e.g. general rule to specific cases; concepts or facts to reasons, theories to opposing theories; etc.

There is no limit, no maximum to deep learning.

Deep and MCQs

I heard a talk from an Engineer saying that, worried by the fact that with MCQs a guess will be right, say, 25% of the time, he tried testing every concept by 3 isomorphic questions: different words and structure but exactly the same concept being tested.

Not only did this reduce chance correctness from 25% to < 2%, but when he interviewed students the ones how had not understood the concept well did not even recognise that those 3 questions had been related.

So MCQs may be better at deep testing than you might realise.

Part 4:

You want understanding? Just test for reasons

MCQs that ask for reasons not truths

Simply asking for reasons, not facts / reproduction of statements, prompts deeper learning (i.e. learning to connect conclusions to supporting reasons).

Assertion-reason questions

Assertion-reason questions

Asking not about the truth of facts, but the reasons for a true fact.
- The question states a fact
- The response options list reasons
- All of these may be true
- All should have been in the course being tested
- => Then recognition will not help the student.

CAAC (Computer Assisted Assessment Centre) website advice on MCQ design:
http://www.caacentre.ac.uk/resources/objective_tests/index.shtml

Questions about reasons (sky)

The night sky is dark because:

A. The Universe is infinite in size

B. The Universe is expanding

C. The Universe is made up of, mostly, Dark Matter

D. The Universe has existed for only a finite amount of time

E. The distribution of stars is not smooth, but fractal

19

Similarly ….

If you want to test for connections to reasons, use MCQs that ask for reasons.

If you want to test for connections to concrete cases, use MCQs that ask about concrete cases. (cf. brain teaser questions)

If you want to test for connections to counter arguments, use MCQs that ask for counter arguments.

20

Part 5:

Aropä and Reciprocal Peer Critiquing (RPC)

Learners being asked to judge each others' work (to exercise "evaluative judgement") prompts (deep) learning.

21

My current recipe for RPC

Reciprocal Peer Critiquing (RPC)

Psychology level 3 undergraduates.

Done twice, first with past (already marked) work;
second for new coursework before submission.

- Students bring in and exchange work
- Prefaced by 1-3 questions they particularly want comments on
- Each critiques 2 others, address criteria plus the questions; rubric: best and worst feature
- Round table, F2F feedback, tutor chairing

22

My current recipe (2)

Always goes down well with my students, once they've done it.

See Morrow (2006) for evidence.

Most enthusiastic about seeing how other students write, but also about getting feedback.

Perhaps best indicator is that having done it the first time, they commit to finishing the next bit of work a week early to allow time to do it then.

23

Prompt sheet

Criterion 1: quality of literature research
What's good?
What could be improved?

Criterion 2: quality of the write-up
i.e. well presented and clearly structured?
What's good?
What could be improved?

Criterion 3: quality of Critical analysis
What's good?
What could be improved?

Prompt sheet 2

This rubric was for an English course:

What is the issue that the draft is addressing? Is it interesting, or do you care?

Say what you think is the argument of the draft. If the argument is not clear, suggest what a possible argument might be.

What reasons does the writer offer to support the argument? (You may like to break down the argument into quasi-syllogistic premises or to identify a Toulmin-style warrant for the argument).

Suggest a counterargument to the argument of the draft. This comment may, alternatively, point out unexamined assumptions and/or missing or unacknowledged evidence.

Identify a characteristic sentence of the writer. Say what you think is good about this sentence, or how this sentence can be improved (your chosen sentence may simply identify a repeated writing fault)

Reciprocal peer critiquing: boxes ticked

Boxes ticked = principles enacted:

- Peer assessment (the peer voice)
- Exercise the criteria from another viewpoint
- Peers see each others' work (resource for remedies)
- See how own and others' work compares in quality
- Learners proactive in formulating feedback questions
- Can act on feedback directly (in 2nd application)
- F2F delivery means dialogue around feedback, and not just clarification but multi-party discussion.
- Multiple opinions on same work: information on variability
- Teacher scaffolds first RPC, then leaves it to the learners

Big scale RPC

What about big classes?
As described, it works for groups of 2-6.

1. I've done it in a lecture group of 90 for short (100-200 word) passages: swap with neighbour and do RCP

2. Use software to manage it. (Aropä)
There is free software, and numerous papers reporting experience, on how to do it with big classes (60, 600, ..)
John Hamer: google "Aropa peer"

27

Anonymous vs. F2F feedback

Pro-anonymous: data protection, privacy

Pro-face to face:

- More useful and serious critiques are elicited
- Dialogue for clarification of what the feedback means
- Dialogue of a more open-ended and multi-party kind
- Get feedback on the feedback you gave
- Hear critical issues directed to others but relevant to self. I.e. discussion of other work than only your own.

28

Part 6:

Mazur's "peer instruction": Disagreeing over the right answer

When peers disagree with each other over what is true, this prompts ways to work out how to judge: generally, by producing reasons for an answer in order to judge amongst them.

29

A widely successful method

There is now a widely successful method, based on replacing lectures by a series of carefully designed MCQs used in big classes.
These are answered by students, but the answers are not given.
Instead, students debate with their neighbours about which is the right answer.

Most often called "peer instruction" now.

Page 30 of 56

Hake's survey

Hake (1998) published a survey of 62 courses (6,542 students) all studying the same subject, all using the same standardised test, and using it both pre- and post-.

He graphed the mean gain on each course against whether or not it had used the method of "Interactive engagement".

And showed a nearly complete superiority of "IE"

31

See fig. 1 in:

Hake,R.R. (1998) Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses Am.J.Physics 66(1), 64-74

Hake's results

32

Mazur

Crouch & Mazur (2001) published an analysis of 10 years of Mazur's Harvard course.

Again, the standardised pre- and post-test.

He concluded he had doubled the amount of learning, but the graph suggests that really, he tripled it.

33

See fig.2 in:

Crouch, C.H. and Mazur, E. (2001), "Peer Instruction: Ten years of experience and results" American Journal of Physics 69, 970-977

Mazur's gains

34

The learning design

So more effective teaching can be achieved.

"Interactive engagement" and "peer instruction" revolve around asking students questions. These may be presented using Electronic Voting Systems (EVS).

But what kind of questions? Brain teasers.

35

Brain teaser questions

The point is to provoke debate, internal and between peers.
Cf. Socratic questioning, and "catalytic assessment"

Remember the old logo or advert for Levi's jeans that showed a pair of jeans being pulled apart by two teams of mules pulling in opposite directions. If one of the mule teams was sent away, and their leg of the jeans tied to a big tree instead, would the force (tension) in the jeans be:

- half
- the same
- or twice what it was with two mule teams?

36

Peer Instruction: Mazur Sequence

1. Concept question posed (brain teaser)
2. *Individual Thinking*: students given time to think individually (1-2 minutes)
3. Students provide individual responses
4. Students receive feedback – poll of responses presented as histogram display
5. *Peer Discussion*: students instructed to convince their neighbours that they have the right answer.
6. Retesting of same concept
7. Students provide individual responses (revised answer)
8. Students receive feedback – poll of responses presented as histogram display
9. Lecturer summarises and explains 'correct' response

37

Evidence that PI causes deep learning

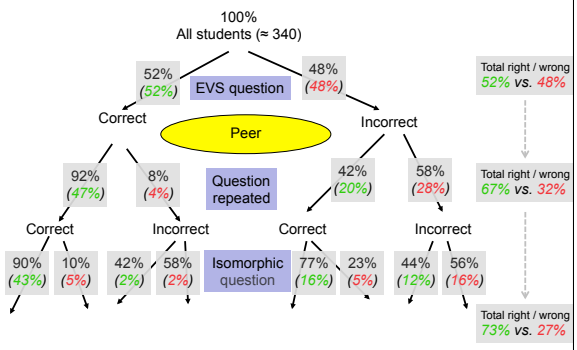
Smith et al. 2009 in *Science*

This paper not only shows that the method has spread across many institutions, but has also transferred from physics to genetics.

The study reported was overtly to test a criticism of peer instruction, that many learners might be learning the right answer from peers, without understanding the principles: i.e. surface not deep learning (when the whole motivation of the method had been to do the reverse, and get students to understand the issues and not just learn to pass exams).

The data not only support this, but also show it isn't fool proof (all students get better), but nevertheless shows a powerful tendency.

Smith et al. 2009 paper in Science v.6



Why "peer instruction" – brain teasers – work [Pillar 2]

Peers are better than experts in some key ways.

When peers disagree with each other over what is true, this prompts ways to work out how to judge: generally, by producing reasons for an answer in order to judge amongst them.

They aren't asked for reasons: it is socially natural to produce and debate them.

See Draper (2009) "catalytic" paper for why and how this type of peer interaction is successful at promoting learning.

Page 40 of 56

Pillar 3: Peer interaction types

There are 2 types of peer interaction:
- Where the group has a joint product to create
- Where there is no product, just a loosely shared topic

In academic conversation there is no requirement to come to an agreement BUT we are very often benefitted by the other acting as a prompt to perturb our private thinking.

41

Part 7:

Learners designing MCQs: Learner-authored questions PeerWise

Teaching someone else is another classic and powerful way of noticing you don't understand something you thought you did, and to fix it.

Learner authored questions

This is another powerful teaching tactic.

Basic idea:
Students have to design a test MCQ (best done by small groups) complete with reasons why each response option is right or wrong (which can then be used for automatic feedback for those attempting the question).

Have to aim for questions that discriminate (splits class).

Why is this effective? The factual question requires them to generate reasons ….

43

Learner authored questions (2)

This has been done numerous times, if you search the literature.

- In multiple disciplines e.g. accountancy, ...
- For credit, or not.
- With a promise to use the best MCQs written by students in the final exam, or not.

For more references see Draper (2009), and Nicol (2007)

44

MCQs as learning objectives

Given that the exam is to be about learners giving reasons, not facts, you could set the work using MCQs as learning objectives.

The learners must then research what the right answers are, and why.

Jaye Richards did this in a school class: a "Jigsaw" design, where the learners, not the teacher, did the exposition.

45

PeerWise

There is a free software service that allows you to do this on any scale, including huge classes. (Currently, its web page says about 200 HEIs (worldwide) have an account with them.)

Students create MCQs and add them to a bank, where their whole class can use them. There are mechanisms for student reviews of how good each question is.

http://peerwise.cs.auckland.ac.nz/
The central website

http://www.psy.gla.ac.uk/~steve/localed/innovs.html#pw
Short introduction to it by me

46

PeerWise (2)

It will only work well if you make the authoring a compulsory assignment (unless you have a class of 1,000; in which case the 5% of volunteers will produce a critical mass of questions the whole class benefits from).

Could make them author MCQs as part of EVS-enhanced presentations they give to the rest of the class.

They will need a bit on instruction on what makes a good question i.e. not too hard or too easy.

PeerWise has systems for students to rate other students' questions, as well as to answer them.

47

Teaching as a prompt to deeper learning

Teaching someone else is another classic and powerful way of noticing you don't understand something you thought you did, and to fix it.

Summary

The 3 "pillars"

1: [Hunt] Realising we aren't sure is a trigger to learning.

2: Peers are particularly good as a trigger / catalyst / prompt.

3: The two types of peer interaction.
And how the learning which the academic no-joint-product type can trigger is not transmission nor pooling of information.

Prompts to deep learning, to understanding

- The need for more **deep learning** i.e. understanding, linking.
- Solo catalysts:
  - Stopping to gauge one's confidence in knowing
  - Answering questions about which is the right reason
  - **Formative tests**: how regular objective tests (e.g. MCQs).
- **Peer interaction**: a powerful prompt to learning. By:
  - Teaching someone else: Learners designing MCQs
  - Critiquing: judging peers' work
  - Discussing which answer (which peer) is right. (brain teasers)

Final joke

Is shallow learning a special human adaptation (other animals, it is currently thought, seem to imitate ends not means)?

And if so, then why is it functional?

A place to stop

For the slides, handout etc. see:

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

xxx

x ➔