

Applied psychology: Education

Steve Draper, Glasgow University

References:

<http://www.psy.gla.ac.uk/~steve/courses/cereRefs.html>

L3 Designated4 "CIP" 14 Oct 2013

1

Introduction

My research interest is "education": specifically,

- 1) How does the learning and teaching process work in higher education (HE);
- 3) And what methods (= interventions = learning activities = learning designs) actually significantly improve learning.

In two hours, I can only pick a few interesting things out of the pot. Perhaps most interesting, yet most often neglected, is some really big and surprising effects. What they mean is not necessarily clear

2

Part 1:

Some mind-boggling effects

3

A fundamental question

A key reason for focussing on big effects is as a way of asking the question: "What difference, if any, do teachers make?"

Or do learners in HE learn regardless?

What use are teachers (lecturers, professors, ...)?

4

Effect no.1: Learning does not need teachers

Allen Tough, a Canadian researcher active in the 1970s, (<http://allentough.com/>) looked into:

How much learning adults did:

- 90% had done at least one project in the last year
- Average 5 projects per year
- Average hours per week: 10

How much of it was independent of courses and teachers (about 4 out of 5 projects).

But also notable is that almost no-one at first said this: they actually didn't realise that this was serious learning, and largely self-directed and self-managed.

5

Allen Tough's Adult learning projects (2)

Definition: any period of time in which your primary motivation (over 50% of the motive) is to gain and retain knowledge and skill.

(N.B. very many of these are for practical reasons, but you pursue the eventual practical end through spending time first directed at learning.)

Spend a few minutes writing this down (then we'll do some sharing):

How many such projects have you done in the last year?
Start writing down the ones you can remember; and if possible, a guesstimate at how many hours altogether you spent at it.

6

Allen Tough's Adult learning projects (3)

Tough and his followers have found a similar pattern in samples from age 16 to 60.

Clearly people have no trouble doing learning, nor in managing their own learning, and more often than not do not find organised teaching (courses) the most useful for their purposes. LifeLong Learning is not new, and doesn't seem to need help.

Most of the literature turns its back on this; Just as drug companies might prefer you not to compare the effect of their drug with the effect of giving no treatment at all.

7

Effect no.2: General effect of teachers

According to Dylan Wiliam:

The biggest causal factor in most studies, is which teacher a child gets: differences (in learning outcomes) are almost always more affected by which teacher than by whether you get the "new" or old teaching method in an experiment.

It will make more difference to a child whether they get the best or worst teacher in a given school, than whether they go to the richest or most "deprived" school in a region.

Some of this may be how many years experience the teacher has: but perhaps only the first year matters much.

8

Implications: effect of teachers (2)

Even if teachers get better by practising, there is no way to give every child a good teacher.

Important aspects of teaching must be, not conscious knowledge, but a tacit skill.

It implies that teacher training may be useless: If we knew what good teaching was, then surely we could teach it to teachers and avoid the dependence on experience?

When we do experiments on teaching methods, we should use as a scale, the size of change that the teacher makes (i.e. the difference between the best and worst teacher). Methods with as big an effect are important....

9

3 roles of teaching

Even if teachers increase learning, there is the question of what it is that they do which matters for this.

(Academic jobs are typically expressed as having 3 kinds of work: Research, teaching, administration.)

But in fact, teaching has 3 facets. And a person might be excellent at one, yet rubbish at another. I.e. good teaching is not a single thing

- Delivery e.g. lecturing, facilitating discussions.
- Knowledge selection and expression.
Selection of topics; selection or authoring of materials
- Designing learning activities.

10

Role 1 of teaching: Delivery. The "Dr.Fox" experiments

There have been experiments on whether lecture delivery skill made a difference to learning.

Hired an actor; fixed the script; had it delivered with high or low "expressiveness".

But also, tried it on 2 groups: students who expected to be tested, and students who did not.

If they thought they had to learn it, it made a difference to their ratings of the lecture but not to their learning (test scores); If they thought they didn't have to learn it, then the well delivered lecture caused higher learning.

So in HE, student will power overrides teacher delivery lack of skill.

11

Role 2 of teaching: Creating knowledge. Chick sexing

To be learned: sexing day old chicks (for the egg industry)

Viewed as an implicit skill: some people could do it, but couldn't tell you how to do it yourself.

Training used to take 6-12 weeks to get a person up to speed and accuracy for useful employment. Method was loads of practice, feedback from an expert.

Then researchers worked on creating an instruction leaflet (pictures, some text). Trainees learned more in 1 minute from the leaflet than in previous 6-12 weeks.

This is an improvement of about 26,000 times.

So discovering the knowledge, and expressing it in a leaflet can be very valuable: one role of a "teacher".

12

Role 3 of teaching: Learning activity design

Mastery Learning

Hake's "Interactive engagement"

Mazur's "Peer instruction"
(really, peer discussion of brain teasers)

13

Effect no.3: Mastery Learning has a huge effect

According to Bloom (1984), Mastery Learning improves learning by 1 standard deviation i.e. an effect size of 1.

14

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.

15

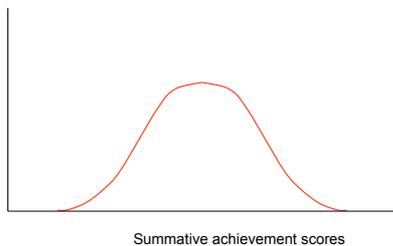
Carroll's view of the determinants of learning

J.B. Carroll's view, backed by experiment, was that learning outcomes depended on all 5 of these factors i.e. is a function of 5 variables (not on a single learner attribute):

1. Time allowed
2. Perseverance
3. Aptitude
4. Quality of instruction
5. Learner's ability to understand the instruction

16

Typical test scores



17

Mastery learning scores



18

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 different stance on capabilities.

19

Mastery Learning: the method

Mastery Learning demonstrated the same effects as the recent published studies mentioned above, more widely, 25 years earlier, using similar methods.

- Telling the students not to interpret formative tests as ability measures
- Giving them highly specific suggestions about how to improve, and the occasions to act on this.
- Showing confidence in them, based on most of the class succeeding
- Giving them the experience of success on objective tests
- I.e. basing assertions on evidence not empty words

One difference was that the first and original aim was to change the mindset teachers have about learners: to convince them that almost all learners can succeed, and that exams are NOT there to label student performance as a measure of capability.

20

Effect no.3: Mastery Learning has a huge effect

According to Bloom (1984), Mastery Learning improves learning by 1 standard deviation i.e. an effect size of 1.

21

Effect no.4:

Brain teaser questions and peer discussion

A method teaching first year physics using interactive discussion in class between peers, stimulated by special "brain teaser" questions has a large and widespread effect.

Hake called the method "interactive engagement".

Mazur called the method "peer instruction".

22

Hake

Hake (1991): "The results [course feedback] showed quite clearly that my brilliant lectures and exciting demonstrations on Newtonian mechanics had passed through the students' minds leaving no measurable trace. To make matters worse, in a student evaluation given shortly after the exam, some students rated me as among the worst instructors they had ever experienced at our university. Knowing something of the teaching effectiveness of my colleagues, I was severely shaken."

So he went looking for better ways to teach physics

23

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".

24

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

25

Mazur

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

Again, the standardised pre- and post-test.

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

26

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

27

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

28

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?

29

Smith et al. 2009 paper in Science

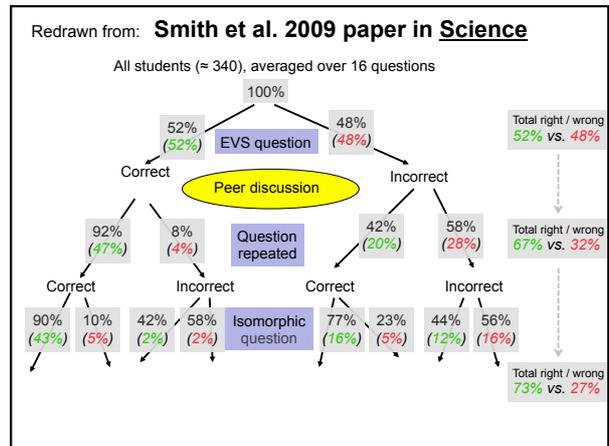
- IE / Mazur type method, but in level 1 Genetics, not physics
- Re-test was not only the identical question, but also another similar (isomorphic) one.
- Even when no-one knew the right answer, many students learned from the peer discussion (for 15 of 16 topics)
- Biggest improvement on the more difficult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



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.)

33

Part 2: An explanation for Hake's / Mazur's success: 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.

Key paper: Draper 2009a

34

Asking about confidence

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 general issue is sometimes called "metacognition": when the learner isn't just a recorder of information but reflects on their learning and may modify their learning activity because of this. Cf. "metamemory".)

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

35

Miyake and "constructive interaction"

We can understand Hake's and Mazur's demonstrated practical educational successes in terms of the theory developed in developmental psychology of how peer interaction promotes individual's conceptual advances.

Miyake (1986) got researchers round her lab to discuss their understanding of sewing machines.

Detailed analysis of the conversations showed that this was NOT teaching, yet both did advance their conceptions.

36

Christine Howe's work (1)

Long series of studies on peer interaction causing conceptual development.

Good selected paper:

Howe, C.J., Tolmie, A, and Rogers,C. (1992)

To get the effect, you need to work on the setup:

Peers with different prior beliefs

Elicit commitment to their personal view in advance e.g. write their view, then show peers this opinion.

37

Christine Howe's work (2)

- Benefit is delayed (e.g. 4 weeks)
- Final conceptions are different in solo than group interviews
- More advanced child ALSO advances still further
I.e. it is NOT information transmission
- "not agreement but private conflict resolution"
⇒ Mechanism is metacognition
(Howe, McWilliam, Cross 2005)

38

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.

Good questions provoke deep(er) learning.
(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

But it appears that who asks the questions can have a big effect too: Peers. Oneself.

39

Two kinds of group interaction

In the "catalytic" paper I suggest that the educational benefits demonstrated by Hake and Mazur depend on the peer discussion in class; and that that process is explicated by Miyake and Howe's work.

Another important implication of their work is that there are two apparently similar but actually quite different kinds of group interaction.

Their kind is CI (constructive interaction), where there is no joint product required of the group: the benefit is reciprocal but private and different. It is useful whether or not they agree. The more often noticed kind of group is in contrast organised to produce a single product: for them, agreement about action is essential, but they generally do not share much knowledge.

40

Two kinds of group interaction (2)

Generally speaking, education requires CI groups (and fails to provide them except by accident);

Work places generally require joint-product groups. Most "groupwork" in universities therefore may not be much preparation for work.

Do not forget that there are BIG penalties for group work. A lot of effort goes into coordination.

In computer programming, one published estimate was that 2 good programmers collaborating produce only 1.1 times as much as a single programmer.

41

Part 3:

Effect size in applied work

42

Effect size

Effect size is a simple statistic (the ratio of <the difference in the means> to <the standard deviation>) indicating how big a difference there is between the two groups compared to the amount the groups vary within themselves in attribute measured.

For pure research, any difference whatsoever, if it is not a chance effect, is equally interesting.
For applied research, only big effect sizes are important.
Why waste effort on something with tiny effects on learners?

The effects I've covered so far were selected for having big effects.

43

NewSci on gender diffs

From: New Sci, 8 March 2011 "Boy brain, girl brain"

TRAIT	Effect size
Gender identity	11.0 - 13.2
Sexual orientation	6.0 - 7.0
Preference for boy's toys	2.1
Height	2
Preference for girl's toys	1.8
Physical aggression	0.4 - 1.3
Empathy	0.3 - 1.3
Fine motor skills	0.5 - 0.6
Mental rotation	0.3 - 0.9
Assertiveness	0.2 - 0.8

Bloom's 2 sigma paper (1984)

When I finally, rather recently, came across an education paper that was reasoning like this, I woke up. His outline argument:

- A) Proof of attainable benefit: skilled 1:1 (or 1:3) tutoring showed a 2 sigma (standard deviation) improvement relative to a school classroom.
- B) But the cost is unaffordable for wide applicability
- C) Bloom's Mastery Learning (ML) gets 1 sigma improvement
- D) => he asks what additional intervention could get the 2nd SD?

Sigma == standard deviation == "effect size"
(effect size = the diff. in the means divided by the SD of the control group)

2 SDs means the *average* tutored learner achieved more than 98% of standard classroom learners.

(Normal distb.: 68.2% within +/- 1 SD of mean; 15.8% above 1 SD; 2.2% above 2 SD.)

46

In search of what makes a difference

Much of the time I've spent looking at theories of the LTP (Learning and Teaching Process). Now I'm trying to apply the Unix rule and Bloom's paradigm for educational research seriously. Where are the big gains to be had?

To advance theory, finding unexpected effects and working to explain them is the way to go. For example ...

An example of a non-priority

The mind-boggling papers showing that where a student sits in a lecture theatre causes a difference in their final course mark. Perkins & Wieman (2005). Random assignment of seat position. Mean final grade of each group "on the edge of significance" but bigger effect on top 20% and bottom 10% of marks; and on attendance.

Griffith (1921) is similar. About 10% mark difference between back and middle of the room.

(But Kalinowski & Taper, 2007 found no such effect.)

This is amazing, worrying, and interesting. But the effects are not very big: this is not a bottleneck, significant as it may be for theory.

47

In search of what makes a difference (2)

So: never mind advancing theory: what is most likely to make a difference, rather than to describe and explain what goes on?

So a first clue is to scan the literature for the biggest demonstrated effects (improvements in learning outcomes). The bulk of this talk is a tour of some truly big effects in education, starting with the least and working up.

But first: do teachers even make any difference anyway?

48

Part 4:

The effect of Expectations; Magic Rubrics

49

Expectations

- Basic effect (punch card training; my microCT exercise)
- Rosenthal's Pygmalion effect of Teacher expectations
- Draper 2009b paper: an interpretation of learners' self-adjusting decisions (including expectations)
- Dweck (Mueller & Dweck 1998)
- Stereotype threat

(see <http://www.psy.gla.ac.uk/~steve/located/dweck.html>)

50

Mueller & Dweck 1998

10 year old US school children, Raven's matrices

1. Medium difficulty set or problems
2. Marks plus praise for a) nothing or b) effort or c) ability
3. Harder task set
4. Told they had done badly
5. Medium difficulty task set
6. Results: a) ± 0 b) +10% c) -10%

N.B. intervention 5 seconds (one sentence)

Effect on task performance rather than learning (strictly speaking)

51

Mueller & Dweck 1998 (text)

"Wow, you did very well on these problems. You got [number of problems] right. That's a really high score."

Either: nothing.

Or: "You must have worked hard at these problems."

Or: "You must be smart at these problems."

52

Dar-Nimrod et al 2006

Canadian female undergraduates

1. Maths test
2. Reading comprehension test (with intervention material)
 - a) material argued that no real gender differences in maths
 - b) material argued there are real gender differences in maths
3. Maths test

Significantly better performance on 2nd maths test if (a): scored \approx 20% better than (b)

Effect on task performance rather than learning (strictly speaking)

53

Cohen et al 1999

US undergraduates, AA (African American) and EA
Write a letter of commendation for publication for a teacher they have known.

Week 2 they get feedback on their letters, extensive criticisms and remedial actions to take. Preceded by a) no rubric b) praise rubric c) magic rubric
Pre/post measures of motivation, and of self-rating at writing skills.

Motivation: if AA and (a), drop in motivation score
Only if (c), higher attitude on writing skills for both AA and EA

54

Cohen 1999 (text)

"It's obvious to me that you've taken your task seriously and I'm going to do likewise by giving you some straightforward, honest feedback. The letter itself is okay as far as it goes—you've followed the instructions, listed your teacher's merits, given evidence in support of them, and importantly, produced an articulate letter. On the other hand, judged by a higher standard, the one that really counts, that is, whether your letter will be publishable in our journal, I have serious reservations. The comments I provide in the following pages are quite critical but I hope helpful. Remember, I wouldn't go to the trouble of giving you this feedback if I didn't think, based on what I've read in your letter, that you are capable."

55

Cohen et al 2006

US middle school children,
Early in the year, a 15 min. written assignment
Choose and write about your (a) most or (b) least important value

End of semester: if African American and (a) then +0.25 of a grade

This reduced failure rate (< D grade) from 20% to 9%

56

Simpson et al 2008

Open University undergraduates just pre-entry

- a) No phone call
- b) A phone call with a script eliciting discussion of student's strengths at learning.

If (b) then -5% dropout in following year (first year)

57

Part 5:

The essay

58

The essay question

Go to this web page:

<http://www.psy.gla.ac.uk/~steve/localed/dweck.html>

You will find several studies both summarised and with references to the published papers.

"From an academic psychology viewpoint, these papers seem unconnected. They each assert a different theory as the basis for the effect they found and reported, and don't cite each other. But educationally, they seem similar in practical importance, in the kind of thing you would do as a teacher, and so on.

Which of these attitudes is right? Is it just a coincidence?"

59

The essay question

It is asking you to discuss how these effects seem similar, yet the psychological theories offered in the published papers seem quite different from each other. Are the theories correct and the similarity of the phenomena just a big coincidence? Or is there a psychological theory that might explain them all, which the researchers have missed?

Starter reference: go to this web page:

<http://www.psy.gla.ac.uk/~steve/localed/dweck.html>

You will find several studies at least both summarised and with the published papers referenced there.

60

The essay question (2)

What do I want?

Well: from an academic psychology viewpoint, these papers seem unconnected. They each assert a different theory as the basis for the effect they found and reported, and don't cite each other. But educationally, they seem similar in practical importance, in the kind of thing you would do as a teacher, and so on.

Which of these attitudes is right? Is it just a coincidence?

61

A place to stop

References:

<http://www.psy.gla.ac.uk/~steve/courses/cereRefs.html>

62