Assessment and feedback: how to get more with less

Steve Draper, Glasgow University

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

Part A:

Introduction: Outline

Outline (1)

This "workshop" is really a trial presentation of a new argument which I expect to be contentious. I will pause after some of the major steps, and invite argument about the soundness of that step.

If it all seems to hold up in your view, then at the end I'll sketch the practical actions you might consider taking as a consequence.

The conclusion I'm coming to is that feedback effort on a course should NOT be evenly distributed (as you would expect if you believe that feedback is generally necessary for learning), BUT should be extremely selectively targeted on:

a) Procedural knowledge (not declarative, conceptual knowledge)
b) Core disciplinary assessment criteria (CDC) e.g. critical thinking

Outline (2)

The major steps in my argument are:

- My department's **anomalous NSS result**: 5th best in the country on overall satisfaction, 101st on feedback.
- Feedback is only important for **learning procedures** not for learning declarative material (facts and concepts).
- **feedForward** is as important as, or perhaps much more important than, feedBack.
  (So having no feedback at all is fine if you have feedforward.)
- Most of the value of feedback is in helping learners to understand **key assessment criteria** e.g. critical thinking. These difficult criteria are core to, and definitive of, the specific discipline. (CDCs)
- **My dept.'s course** gives little feedback, but much feedforward, focussed on procedures, and specifically on CDCs.
- **A 2nd course design** in another discipline has outstanding results, appears different, but actually has similar features.

Discussion?

Any immediate questions about what my argument is at the level of an outline?

And that, if correct, then it would have important implications for the allocation of our effort to assessment and feedback?

OK: then you know where I'm going, and that if it were true then it would be important to you.

Part B:

The nature of theory

Types of "theory" in the area of assessment and feedback (A&F)

- Surface tips e.g. "get marked work back ASAP", "be kind"
- Design principles e.g. Nicol's, Chickering & Gamson's
- Deeper theory (this paper is a fairly rare attempt at this)

The trouble with surface tips is that they summarise a type of student complaint, but are typically only true in some contexts, not others.

Design principles point one level deeper, but don't tell you directly how to generate a great design. Instead, they work by "box ticking": if you come up with a design, then you can estimate its merit by how many boxes/principles it ticks simultaneously.
E.g. "RPC" (reciprocal peer critiquing)

7

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

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?

Nicol's 5/7

Good feedback practice (by teachers to learners on their work):

√ 1. helps clarify what good performance is (goals, criteria, expected standards);
√ 2. facilitates the development of self-assessment (reflection) in learning;
√ 3. delivers high quality information to students about their learning;
√ 4. encourages teacher and peer dialogue around learning;
<This could be improved further>
5. encourages positive motivational beliefs and self-esteem;
√ 6. provides opportunities to close the gap between current and desired performance;
7. provides information to teachers that can be used to help shape the teaching.

David Nicol & Debra Macfarlane-Dick (2006) "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice" Studies in Higher Education vol.31 no.2 pp.199-218

Gibbs' extra 2/4

Assessment activities support student learning if they:

1. Capture sufficient study time and effort (in & out of class)
2. Are spread evenly across topics and weeks
√3. Lead to productive learning activity (deep rather than surface learning)
√ 4. Communicate clear and high expectations.

Gibbs, G and Simpson, C. (2004) "Conditions under which assessment supports students' learning" Learning and Teaching in Higher Education vol.1 pp.3-31.

Chickering & Gamson 4/7

1. Encourages contacts between students and faculty.
√ 2. Develops reciprocity & cooperation among students.
√ 3. Uses active learning techniques.
√ 4. Gives prompt feedback.
5. Emphasizes time on task.
6. Communicates high expectations.
√ 7. Respects diverse talents and ways of learning.

Arthur W. Chickering and Zelda F. Gamson (1987) "Seven principles for good practice in undergraduate education" American Association of Higher Education Bulletin pp.3-7

NSS A&F subscale (5 items) 4/5

UK National Student Survey:
its subscale on assessment and feedback.

√ 5. The criteria used in marking have been clear in advance.
6. Assessment arrangements and marking have been fair
√ 7. Feedback on my work has been prompt
√ 8. I have received detailed comments on my work
√ 9. Feedback on my work has helped me clarify things I did not understand.

13

Rowntree 10/17

Rowntree's book ends with 17 recommendations.
RPC is neutral for 6, ticks 9, excels at 1, and contradicts 1.

X 10. Report feedback/marks only to learner (not in public)

√√ 7. Markers may have quite different perceptions. Report divergence

It ticks 1,5,6,7,9,11,12,13,14,15, including:
1. Articulate the assessment criteria
5. Give learners feedback on their qualities as well as on the product
9. Support portfolios, that include both products and judgments
11. Emphasise learners' strengths, but mention weaknesses
12. Multi-dimensional marks, not portmanteau grades

Beyond box ticking

Surface tips prescribe surface actions like "fast turnaround", "don't give more than 3 bits of commentary at a time"

Design principles (e.g. Nicol's) prescribe more general classes of action e.g. use peer feedback, focus on assessment criteria

The argument in this workshop attempts something much more precise: that would in effect tell us when we needn't bother adhering to those other principles and tips, and when they would have big effects.

15

Part C:

Anomalous NSS 09 results

16

Anomalous NSS result

The last (2009) NSS results threw up a striking anomaly for my dept.
There is one overall question (22) on satisfaction: this is how departments rank themselves relative to others
There are 21 other questions, which implicitly should indicate areas that contribute to the overall result.

In the numbers following I made a number of choices: the basic argument is not sensitive to changes in these assumptions.
Ignore part-time degree returns (else you get several entries per HEI)
Used measure "average response", not "proportion satisfied"

17

Ranks

Qu	score	Percentage agree	Rank	Rank % agree of 107
14	4.60	95	1	*3*
15	4.54	96	2	*3*
22	4.44	96	5	*3*
11	4.43	97	8	*3*
6	4.15	90	8	*5*
16	4.48	91	8	*16*
13	4.48	93	11	*14*
1	4.22	98	16	*8*
4	4.36	94	16	*12*
10	3.96	79	21	*19*
2	4.07	90	22	*16*
18	4.31	85	22	*35*
3	4.26	89	25	*22*
12	3.90	74	31	*27*
5	3.91	72	35	*47*
17	4.30	91	40	*28*
21	3.96	77	46	*36*
20	4.06	78	48	*48*
19	3.90	76	51	*41*
7	3.37	52	54	*61*
9	3.12	42	79	*82*
8	2.83	33	101	*104*

18

Qu	Rank	Qu.text
1	14	Any changes in the course or teaching have been communicated effectively.
2	15	The course is well organised and is running smoothly.
5	22	Overall, I am satisfied with the quality of the course.
8	6	Assessment arrangements and marking have been fair.
8	11	I have been able to contact staff when I needed to.
8	16	The library resources and services are good enough for my needs.
11	13	The timetable works efficiently as far as my activities are concerned.
16	1	Staff are good at explaining things.
35	5	The criteria used in marking have been clear in advance.
54	7	Feedback on my work has been prompt.
79	9	Feedback on my work has helped me clarify things I did not understand.
101	8	I have received detailed comments on my work.

Anomalous NSS result (2)

Got rank 5 of 107 overall

But got ranks much lower than this for 19 of the 21 questions

How is this possible? Two types of explanation.

A) Radically different weightings for each item's contribution

B) There is a large mystery factor not measured by any NSS item

20

A. Different weightings

For this explanation to work, we have to believe that qu.s 14, 15 carry (much) more weight than anything else including feedback:
Course well organised
Any changes communicated well

This would work if primary exposition by staff or self-teaching by students worked so well that feedback wasn't needed …

21

B. Missing factor(s)

How big a factor?
Look at the gap G between actual mean qus1-21 score, and the one that would be necessary to get same high rank on that basis. G=0.09

a) If it were a secret ingredient, then this dept. would get 5s for it, others would get 1 for it. Ratio of weights for new factor to NSS factors would be G/4 = 0.0225
b) If difference in scores for mystery was like the 1.5 difference of best and worst scores for qu.22, then ratio = G/1.5 = 0.06
c) If difference more like the scores on qu.22 for rank 8 and 17, then ratio = G/0.17 = 0.53. I.e. missing factor needs to be half as powerful as all the NSS items together.

22

B1. Secret ingredient

E.g. Drugs, bribes, flowers in the classroom

I haven't spotted one that we do and others completely omit.

23

B2. Missing factors to do with feedback

Another possibility is that the missing factor is to do with feedback, but missed by NSS questions.
Then it is the last, biggest, ratio of weights we might expect.

We're looking for something others do a bit of, but we do better, and which matters rather a lot (as much as 10 other NSS items combined).

This is the possibility I'm exploring in this talk/workshop.

24

Discussion?

Can you see flaws in my argument on this?

25

Part C-2. The programme design

The anomalous NSS result refers to my own department of psychology, and NSS responses made by students in about January 2009 (in their final year). I will assume that the design of their 3rd and 4th year programme mainly shaped them.

In level 3 they took 9 content modules, and received no feedback or tutorials for any of these. In level 4 they took 6 content modules without feedback. However at the same time they received a mixture of small group (6 students) and individual tutorial guidance, i.e. feedforward, on major coursework projects. Three of these are "critical reviews" which focus on demonstrating critical thinking applied to published literature.
Critical thinking is also a marking criterion for the written exams on the content modules.

26

Part D:

Feedback is more important for procedural knowledge

27

Procedural vs. declarative knowledge

"Declarative" knowledge includes facts, concepts.
"Procedural" knowledge is knowing how to do something.

Learners may need testing on declarative knowledge but in fact they can test themselves; and more importantly, they can self-correct once they have noticed they got a fact or concept wrong. Human feedback is seldom essential, though getting them to **use** material in order to provide occasions for noticing their own gaps is important.
(Catalytic assessment; confidence testing; ….)

28

Procedural knowledge

A fact is a single item: if it's wrong, there's no puzzle where the problem is.
Any procedure, however, is a long sequence of actions. If it's wrong (produces the wrong answer or effect) it is generally not clear which part of it was wrong. Diagnostic feedback is very important, though advanced learners eventually acquire sophisticated meta-procedures for self-diagnosis of failed procedures.
If you give me an essay and I just say there's something wrong with it, that is of little help. If you bake a cake, and when it comes out of the oven it is obviously bad, again you are often at a loss. Similarly if your computer program just fails, you don't know which line and which character is responsible.
Generally speaking, procedures have many more component parts; and learners are much more in need of both practice and helpful feedback in learning them.

29

Feedback on procedural learning

The literature also supports this, that feedback has more positive effect on learning for procedures than for declarative material.

Hattie & Timperley (2007) "The power of feedback" can be interpreted as arguing that feedback on procedural learning is more important, because (only) it leads to transfer.

30

Discussion?

Can you think of evidence of learning declarative knowledge without feedback?

Can you think of evidence of needing feedback when learning procedural knowledge?

31

Part E:

FeedForward not feedback

32

Feedforward

In almost all the educational literature, in everyday understanding (of students answering the NSS), and in engineering, "feedback" means giving a measurement on finished work.

Feedforward means taking a measurement now and estimating what will be required for success in the future. This is what project supervisors do.

PhD students only ever get feedforward (apart from post-viva corrections): they cannot ever get feedback.

Educational feedforward means essentially taking the learner's plans and steps so far, and giving direction on what's good, what should be changed.

Feedforward may be somewhat harder for tutors to provide than feedback, since they cannot see the full effect of a learner's plans until the work is completed. But for the learner, it must be much more useful, and saving of time and effort, because it saves the waste of building on a mistaken step.

33

Feedforward (2)

Another way to look at it is to imagine what inner mental changes an ideal learner would make in response to feedback, and whether these could all be made in response to feedforward.

Consider the simple case of a learner being told they have misspelled "Wiliam" in "Black & Wiliam". They will hopefully change 3 things without further instruction:

- The spelling used in the current piece of work
- Their internal "generator" so they will spell it correctly when writing in future.
- Their internal proof reader or bug detector, so that when reading over their own or others' work, they will detect if it's misspelled.

They can do the last two just as well from feedback and feedforward, but it will save work on the first, the earlier (more "forward") the correction is communicated. If instead of spelling, it is information on the right statistics test to use, or the right structure for a major essay huge amounts of work could be saved.

34

Feedforward (3)

Feedback on assembling parts into wholes;
for late stage contingent tutoring.

Feedforward for early stage contingent tutoring i.e. for major learning.

35

Discussion?

How much feedforward compared to feedback is there in your department? (or was there when you did your degree?)

Is feedforward better than feedback?

Any examples where one is better than the other?

36

Part F:

Core Disciplinary Procedures are what matters

37

Core disciplinary assessment criteria

We know from the feedback literature, especially Sadler 1989, that a key difficulty for students is understanding the meaning of assessment criteria. Classic ineffective feedback is "poor conclusion" or "not critical enough" because exactly what the student doesn't understand is what is not expressed there: the meaning, and its operationalisation, of "good conclusion" or "critical argument".

Not all criteria are difficult.
But the ones that are, are not just poorly communicated. They typically are the ones that lie at the heart of a discipline's tacit definition of itself. In other words, they are the most important thing a student must learn during their degree; and often, staff cannot easily explain them: they may be held as tacit knowledge.

38

Core disciplinary assessment criteria (2)

There is a real sense that the central learning aim of a history degree is to learn to write a history essay.
In psychology, to write a psychology essay.
In physics, to demonstrate analysis, reasoning and calculation like a physicist (not like an accountant, or mathematician, or logician)

So my department's programme design could be redescribed as investing 100% of its tutor time in focussing on equipping the students with the ability to display critical thinking (of the kind a psychologist values). It invented a type of coursework ("critical review") that announces to students what the main point is; it requires them to produce 3 month long pieces of work focussed on it; but also marks their exams with this requirement applied.

It is the hardest thing they must learn; the most important thing; almost all our teaching investment is put into it; and the students rated us 5 out of 107 in the UK.

39

Core disciplinary assessment criteria (3)

So on this account, the key question for each discipline is:
What is the assessment criterion that is closest to meaning: display thinking like a scholar in this discipline?

Many disciplines in HE already have much of their assessment organised around a single standard format that exhibits this thinking style e.g. essays for most Arts and SocialScience subjects (but actually, quite different essay types depending on the discipline), "problem solving" involving calculation i.e. inferential maths in most science and engineering.

The argument here is: Focus the feedback more effectively, not on the assessment format (i.e. not simply do lots of essays or whatever) but on learners grasping the core criterion.

40

Core disciplinary assessment criteria (4)

In my dept. by having major coursework focussed on the most difficult criterion ("being critical") was part of this. And by having them spend 3 months, rather than the 60 minutes of an exam essay, allowed more focus too.

But I've had promising results from a 5 minute version too: I.e. the radical change in time scale can also help focus on the criterion by changing context.
Reciprocal peer critiquing using the "criticality" criterion also may help here: I.e. exercising the same criterion as a reader-critic as well as as an author.

All of this could be done for other criteria e.g. spelling, reading recent not old literature, etc. But you wouldn't expect the same rewards. This is about focus, not universal tips/approaches to feedback.

41

*Discussion?

What is the core disciplinary criterion in YOUR discipline?

42

Part G:

Vicky Gunn course "B&B"

43

Body & Belief

Honours option, History discipline, over two semesters, mixture of level 3 and 4 students.
50% coursework, 50% final unseen exam
One 2-hour class per week

A major redesign was introduced (same staff), which was evidently a huge success.

44

Learner evaluation: quick measures

B&B is the course you most liked or enjoyed in Honours? **7 of 8**

B&B is your most valuable course in Honours **8 of 8**

45

Learner evaluation: quick measures (2)

Votes for best feature of the course [some top equal votes] (plus in parentheses: rated "very important" or "best"):

4 (8) Interesting subject matter

2 (8) Great discussion and group feeling between students

1 (8) The staff teaching it are great

2 (7) Approaches and skills I learned on this course, I'm applying to other courses

1 (7) The methods of teaching / learning kept my attention

0 (3) I saw my work was useful to others (not artificial)

Student summary statements (1)

[Pleasure, interest]

I interviewed all 9 students, and later agreed with each of them a personal statement indicating the varying attitudes.

"Best course of any in my 5 years as an undergraduate, because it felt interesting. Most important attributes were Vicky's enthusiasm for the course, for the student discussion, for the students' work (what they were doing). Also important was the participation of all the students, and the individual written feedback I got". – Nicola

"For me, the pleasure of doing this course is about the mind-expanding ideas. Feeling drawn to doing the reading, enjoying doing the assessments, and finding myself frequently discussing the ideas with people outside the course. An intellectual treat, above all." --Fiona

47

Student summary statements (5) [evidence]

"The course developed my sense of personal mastery: doing the reading myself from original sources, being able (eventually) to understand those things and use them myself. (Being personally original, sharing the reading, enjoying a good group atmosphere were less important for me.)" – Laura

"What I personally appreciated learning from the course, was the help in argument building, which I learned from the excellent, extensive essay feedbacks." –Mira

"While the course content was very interesting, it could never equal that of my beloved Islamic courses. But for real value in developing methods of approaching topics and basing my work on primary sources rather than derivative commentaries, this course was of unequalled importance." – VickyT

Body & Belief (2)

The aim of the redesign was to get the students thinking and working like research colleagues for the staff.
Thinking like a historian.

Superficially very different from the psychology case, it had the same effect of transfer of general core skills (even though it is catalogued as a content option).

49

Body & Belief (3)

The core criterion for this discipline seems, to an outsider, to be something like:

Select one of a set of theorists; apply it to one small period of the overall period of the course; invent a thesis, and argue it using the theorist and that data.
The quality of the argument, not the correctness of the thesis, determine the mark.

N.B. this is critical thinking but is not the same as what (say) a psychologist means by "critical thinking".

50

Body & Belief (4)

They did one early essay;

Then a series of formats on a single topic (each student selected their topic): tutor feedback/forward on their preparations for a presentation; the actual presentation to the class; a writeup of the presentation.

Lots of feedback and feedforward
Linked tasks exercising the same core issues in different contexts;
But also, building on their development of the same (personal) topic in terms of content.

51

B&B as the tutorial component?

The staff:student ratio of 2:9 would seem wildly extravagant to a big department (like psychology).

But as a tutorial group, this is normal: easily justified if it were implicitly carrying the entire tutorial burden of the department. There is no separate tutorial strand in that department.

The fact that students say that B&B has affected their approach to study on the other honours courses (but not vice versa) supports this interpretation.

52

Part H:

Conclusion: recommended actions

53

Strategy?

- Identify your core disciplinary (assessment) criteria (CDC)
- Focus most or all effort on training students on it:
  both student effort and staff effort
- Usually many assessments already do test them
- However the same focus may not be present in the feedback
- Furthermore supplementary exercises may be effective.

54

Why?

- The CDC are the procedural version of threshold concepts.
- They are the hardest things students have to learn
- They are the most important too: almost all assessment in fact uses them.
- They require a longer timescale to master (not one short module)
- The reward is to see this learning transfer across modules; even across years and departments.

55

Tactics?

- Exercise the/each criterion in both directions: not just as authors but as readers/critics
- Exercise the same criterion in tasks that are superficially very different. (learn what is common across contexts)
- Try radically different timescales.
  3 month, 1 hour, 5 minute versions.

56

A place to stop

For the slides, handout etc. see:

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

57