

Assessment in action:

Practical example to implement assessment theory

In the last lecture, we spoke at length about how psychological theory has been used (and abused) to help influence many educational policies and practices when it comes to formally assessing students or pupils.

To make this knowledge clearer and cleaner to understand an examination is designed specifically around this course. This should give an idea as to what the rationale and use of the theory is in a context that you will all encounter.

After I finished the lecture, a number of you (in service) pointed out that you used this format of both designing curriculum content, and also in the assessment that you use for your topic – respect to you. Most people in tertiary teaching that I know of both here and abroad do not. So you're already one up on them in my humble opinion.

Which assessment protocol to use (and why)

If you remember, from the last lecture, there are implied psychological theories that are being used if you assess folks in a certain way. If for instance you assume that the intellectual capabilities are *normally distributed* and that the aim of the assessment is to find the 'top' of the class, regardless of their actual capabilities – then you are buying into a relative assessment tool that does not really care if a student/pupil can actually perform a task. One presumes that the logic is that given this 'raw intelligence' one can subsequently mould it into whatever one wants, the assessment tool is tasked in simply finding the raw talent. The ideal assessment tool is one that provides fine grade discrimination power. That is you don't want to clump all the scores around a few scores, but rather have them spread out over your minimum and maximum scores. In essence this theory is leaning towards the 'nature' or genetic model of intelligence.

If you have a more sociological, or 'nurture' view of intelligence, then you want to have discriminatory power in terms of establishing what a pupil/student can actually do. Either because you want to match vocational skill sets directly, or because you want to establish if students/pupils have identified shortfalls in their own learning - much as Alfred Binet was originally asked to do.

Consideration of the Purpose of an Exam

To be honest, I'm not a fan of examinations. This is something that's already been set and apparently cannot be changed. I believe examinations have the tendency to put us in artificial positions which we're highly unlikely to encounter in 'real' life. An examination has:

- low ecological validity
- low construct validity
- and low criterion validity.

It does however, for mainly historical reasons, have very high face validity. In terms of appearing to be doing the 'right' assessment, an examination fits the bill very well. The standard format would be an examination in three parts: a multiple choice section, a short answer section and an essay section, normally asking for memory of facts, definitions, ability to work through some potential calculations by memorising relationships or formulas. Who would **not** initially think that this is a valid way of assessing anyone?

I did.

Fortunately there are a number of lecturers who are now making their course "examination free". However, we are required to take the exam because this is what I inherited when I came to teach the course. FNU feel that this is a course design element that cannot be changed without going through a number of different committees.

OK, so the question now remains how to make the examination assessment count for something that has good criterion/construct & ecological validity (from now on I'll call all three 'real world validity')?

Norm Referenced Assessment?

Norm referenced assessment does not give me any assurance that my tests have real world validity. The problem is NOT so much in the questions that might be set, but the assumed marking scheme which is overlaid on however good, or bad the questions in the assessment are.

The whole class might answer a question this year exceptionally well, perhaps because by luck they were a very smart class, or by luck they had a great lecturer - or both. Norm referenced assessment would not recognise this. Instead the top 4% achieve the 'A' status, the next 17% the 'B' status and so on. Assuming the discriminatory power was good enough.

Perhaps a worst case scenario, the whole class is really very bad (bad luck?) or perhaps the lecturer is very bad, or perhaps everything is bad and the result is that the students are incapable of teaching ... well anything really. Notice how the norm referenced assessment doesn't really care about this. Norm referenced assessment will continue to dutifully sort the students from 'highest' to 'lowest' scores, never mind that they're all pretty bad. We will be letting a whole year of students loose on a set of students/pupils with qualifications to cause, at best, no real learning; at worst it could really harm the student's/pupil's learning capabilities.

If I'm interested in 'real world validity' then I should definitely **NOT** use norm referenced assessment.

Criterion Assessment?

Since my focus on 'real world' validity, that means that what this assignment measures your ability to have actually learned something on this course. The question now becomes 'how' can I prove to the outside world that if you have learned something from this course, that you can demonstrate this in an objective measurable way? In other words what is the criteria (benchmark) that shows that you've understood the course? Let's start with what the course was supposed to teach you and then work our way forwards into the assessment criteria. In the very first lecture with you, I suggested that if there was 'one' and only 'one' criteria that we should be judged by (yourselves as students and myself as a facilitator/lecturer), it is:

Your ability to use the theories to contribute significantly to being a better teacher/facilitator.

If you weren't able to actually utilise these theories, then to some extent this course has failed you (let's not get into the blame game until that time though). This is the criteria that we're going to continually refer to in order to construct our examination. This criteria was chosen essentially using the *backwards design* approach when T5303 was being designed.

There are some constraints to the format that we're allowed to employ in the exam – again this is not a choice but simply what we have (see the box). I suggested that it would be a number of different essay questions. Here are some questions that were generated in class:

- (i) Write an essay on a psychological theory and how it was used in the classroom.
- (ii) Explain five theories and how it can be used on the classroom.
- (iii) Explain the differences between psychometric theories of intelligence and the *Multiple Intelligence Theory*.

External Constraints

Some constraints we have in our exam (externally imposed)

- * The exam is 3 hours.
- * You need to be mainly communicating on paper, which normally means writing - but hey cartoons, pictures, tables, diagrams, writing in a play format, are all acceptable.
- * Open Book examination.
- * And remember no WMDs allowed in the examination hall!

- (iv) Read a case study and explain which theory is being used.
- (v) State a theory that attracts you and how you would use it in the classroom.

Let's take a bit of time to evaluate the questions.

Explain the differences between psychometric theories of intelligence and the Multiple Intelligence Theory.

Ok, this is never going to be a question, principally because it's a 'definition question. That does not address our criteria of being able to use a theory in a class room setting.

So, let's ditch that.

~~Explain the differences between psychometric theories of intelligence and the Multiple Intelligence Theory.~~

The first two questions are essentially the same question, one asking for a single theory and the other asking for five theories.

Write an essay on a psychological theory and how it was used in the classroom.

or

Explain five theories and how it can be used on the classroom.

The problem with both these questions is that it doesn't directly access the criteria. It's conceivable that a student will simply rote learn a theory and then rote learn the appropriate classroom context in which it can be applied. This would thus provide no evidence that you were able to utilise these theories in a teaching/ learning context.

So, neither of these questions is really suitable and are ditched.

~~Write an essay on a psychological theory and how it was used in the classroom.~~

or

~~Explain five theories and how it can be used on the classroom.~~

That leaves us with the final two questions which do look promising. Let's take the second to last question.

Read a case study and explain which theory is being used.

Yes, this is a recognition of a theory question, so it doesn't quite access the criteria of being able to state how to use the theory in practice. So one of the ways to do this would be to ask also how useful the theory is.

My issue with allowing space for only one theory was that it doesn't allow for a number of theories, especially if they are spread out over 'design' vs. 'curriculum' vs. 'assessment'. If the case study was written so that there really was only 'one' theory it would probably not resemble a real world teaching/ learning contexts whatsoever. I've pledged to write in a number of different implicit theory uses.

The new question now looks like this.

Read a case study and explain which theories are being used, and give an explanation of its usefulness in this educational context.

Dealing the last question which looks deliciously simple.

State a theory that attracts you and how you would use it in the classroom.

So now the questions looks a bit like this

OK, my issue with this question is that potentially you might just answer the question "I like *theory X*, and I'd use it to give everyone a jelly sweet at the end of every lesson". Now strictly you'd have answered the question, so it has to be a bit more than this. I decided to put in that you needed to explain 'why' you liked a theory.

State a theory that attracts you, explain why, and how you would use it in the classroom.

We did some further modifications to this and the first question, principally just tidying up the language and allocating the proportion of marks allocated within each question given in round brackets to reflect the emphasis on how this fits in with the criterion assessment for this course. Notice how the questions were designed specifically to assess the ability of a student to:

... use the theories to contribute significantly to being a better teacher/facilitator.

Even though we know that the exam format is not ideal, it is certainly a better option than having the standard, multiple choice, short answers and essay format which then gets 'norm referenced' anyway.

Final Questions

OK, so at the end of all this we have an exam question that looks like this (proportion of marks within each question in parentheses):

1. Read the following case study¹ on a teaching/learning context. Identify which psychological learning theories are being used implicitly or explicitly (30%). Evaluate if this is an effective teaching/learning strategy and suggest (if any) improvements that might be made to make the teaching/learning more effective (70%).
2. From this course, state a psychological theory (or theories) that has interested you (25%) – perhaps it surprised you, or you felt negatively (repulsed) towards the theory (ies) (25%). Explain your reasons why it is interesting to you. Explain how you will incorporate this information in your future teaching (50%).

Each question is worth 50% of the total exam, ie you should be spending equal amounts of time on each.

¹ Some of you maybe wondering why I don't give you the actual case study right now, so that you have time to consider and reflect on it. Well I've already done a lot of that in your previous three assignments. In this assignment I want to actually test to what extent you can rely on your own resources and materials that you've gathered to make informed decisions. Part of me wants to do that because some decision making, at the very least, has to come from you. As professional teachers, you're not always going to be in a position of always having access to experts, libraries and books - and to be frank it would get too tedious if you kept referring to these sources for 'every' educational decision that you'll have to make.

Grading the Answers

But wait!

Designing the questions is only half the exercise, we still haven't actually operationalised (that is specified which specific 'operations' – a fancy way of saying 'tasks') what it means to pass, or to pass well, or to pass exceptionally well. In other words we're not going to do an assessment task that consists of only two grades (pass or fail) but instead think of five grades.

Rather than state what the grades are and what the points are that achieve a particular grade, it makes sense to consider, in English, what these grades are supposed to represent. So in my book:

- A Grade – means you did exceptionally well, not only did you have all the facts about a topic but you put it together in a unique way that shows not only insight, but also *creativity and lateral thinking*.
- B Grade – means you did really very well. You would have got *all the facts about a topic*, *gained insight* and you were able to *competently* explain the topic, rather than just 'parrot' the facts and figures.
- C Grade – means you were adequate in explaining the facts that shows comprehension of the topic (rather than insight). This is the minimum to pass.
- D Grade – means that you did not have an adequate understanding of the topic. Instead you've started to grasp certain elements of it, and have started to use only some of the appropriate language and terminology to describe the issue under study. However, in sum, it is not enough to be considered 'enough' to be able to state that you've really learned the criteria relevant to this topic. It's a 'failure'.
- E Grade – Means, that you've not really understood any part of the topic. Any marks you've acquired are most likely chance remarks, or individual words and phrases that are relevant. I've suggested that the 'E' should stand for 'echo' of your wailing, after you've been flung off a cliff for doing so poorly at the topic.



Assigning the Criteria to Grades

The thing that we discussed in class, where many of you had suggestions, were not replacing one vague statement with another. That is many of you said something like 'to pass adequate [c grade] would be to do 'ok' in the answer: all this has done is replace 'adequate' with 'ok'.

What follows are the criteria that we assigned to each of the grades. I personally find it easier to start at the 'C' grade which is the 'pass' mark or where students have shown adequate evidence that they've understood the topic. Then you add on top of that to get the higher grades, and subtract from the 'C' grade to get the 'fail' grades operationalised.

Pass/fail	Grade	Criteria for question ① – describe a case study etc. ...
Pass	C	Identify at least one theory correctly, and make a correct evaluation of it's use in the stated context.
	B	As a 'C' grade but can identify other theories too and offer some evaluation as to how it or a competing theory can be used in a better way – or if no better way is possible, then explain why this is so.
	A	As the 'B' grade but has additional unique or imaginative insights in the use of the theory.
Fail	D	Identify at least one theory correctly, but do not supply an evaluation of it's use (or provide one incorrectly).
	E	No evidence that any knowledge of any of the theories or their correct interpretation is used.

Pass/fail	Grade	Criteria for question ② – explain a theory that was/is interesting etc. ...
Pass	C	Can correctly describe the surprising theory & it's correct usage in future teaching.
	B	As a 'C' grade but can also explain the reasons why it is 'interesting'
	A	As the 'B' grade but can give a comprehensive integrated reason why it is interesting along with uniquely generated ideas of the usage of the theory in future teaching that demonstrates genuine insight and understanding.
Fail	D	Describe one theory and give the reasons why it is considered interesting.
	E	No evidence that any knowledge of any psychological theories or why it should be of a professional interest, or indeed how it could be used successfully in future teaching.

I wanted to mention a few things that we were not going to do. One of the class members suggesting that we could employ a few multiple choice questions in one of the sections - err no! Multiple choice essentially tests memorised answers or processes - which rarely have good ecological validity. We also were not going to award a higher grade if someone spelt correctly and used perfect English. The answer was that no this would not count in the formal grade since this is not an English class with English assessments. You are however being assessed on your communication skills. If you can get the message across with cartoon drawings then this will still carry full marks. Having said that for those that do write good English probably have an advantage in communicating clearly – or they do not need to work as hard to achieve good communication.

Summary

Rather than repeat the previous lecture, I thought I'd try and bring out the key points in the assessment lecture to describe how an assessment task is being designed. The actual assessment task will be your 4th assignment, otherwise known as your exam.

We pointed out that the task of this assignment was to have high 'real world' validity, and that we did not require the exam to have good face validity.

Since *real world validity* was of prime importance, we discussed why a norm referenced assessment would not help us to fulfil this requirement. We opted instead for a criterion referenced assessment approach.

I established what the overall criteria for this course was (to be able to successfully utilise psychological theory in present and future teaching) using a 'backwards design' approach. This was our guide to assess a number of class-generated questions that could be asked in the exam. A number of options were dismissed because they could not adequately tell us whether a student had demonstrated that they had passed the main criteria.

We then further refined the remaining questions and loaded their internal loadings to give an adequate reflection with respect to the course criteria.

Finally we established what are operationally designed performance tasks that establish what grade one would get.