"Problem-based Learning: helping your students gain the most from PBL" 3rd edition, March 1996

Instructor's Guide for "Problem-based Learning: how to gain the most from PBL" ISBN 0-9698725-0-X

© copyright 1996, Donald R. Woods

5. Questions and answers about assessment

Most of us are familiar with tests and exams. We have created numerous tests and examinations in the past. We try to create them so that they are "fair" for the students (that is, they are on the subject material we "covered" in lecture), that they can be done (we might try doing the question ourselves and then double the time allowed). We also want them to be easy to "mark." The test should be consistent with our own sense of what we want the students to have learned. Sometimes, our focus is on definitions (or Bloom's taxonomy Level 1); Sometimes we want them to show that they comprehend (Bloom's Level 2) or use the knowledge to solve problems (Bloom's taxonomy Level 3). Bloom's taxonomy has helped us in the past to ensure that we have a range of different types of questions on the "test". Thus, we don't include all definitions; nor do we include all open-ended problems to solve. We also try to include questions that will represent what might be expected of the students later on in their career.

So how do we handle tests in the context of small group, self-directed, self-assessed PBL? In this Chapter, we start with some basics about assessment, list some options and then address a series of questions that are asked often.

5.1 Why is assessment important? Because students learn what is assessed.

"What's going to be on the test?" "Are we going to be tested on this?" These are commonly heard questions from students.

If we don't test it, students won't spend time learning it! If we test by multiple choice tests, then students memorize minute detail and the subtleties of the wording of their texts. If we test by having them solve problems, then they collect as many sample solutions of problems as they can and memorize those prescriptions.

Students study to succeed; they are shown to succeed if they pass tests. They study to pass tests.

The key issue is to assess students so that they learn what we want them to learn; so that they acquire the behaviours we desire as outcomes of the program.

5.2 Background issues about assessment

The overall 16 issues about any assessment (outlined in Chapter 9 of **HTGTM**) are listed and discussed below.

for the expected outcomes

Decide who needs to be involved in the following steps and then:

1.Select the content: general knowledge, attitude, psychomotor or process skills.

2. Express these specifically in terms of written observable goals.

3.Decide who creates the goals.

4.Make the goals explicit and public.

5.Identify measurable criteria that tell the degree to which the goal has been achieved. The criteria **must** be tied to the goals in #2.

6.Decide who creates the criteria.

7.Make the criteria explicit and public.

8.Decide what constitutes evidence to show that the goal was accomplished based on the criteria.

9.Check to be sure that the goals, the criteria and the extent and type of evidence are realistic (that is, can be accomplished with the resources available). If not, **revise**.

the assessment process

10.Decide on purpose of the assessment.

11.Select the performance conditions.

12.Decide who assesses.

13.Set the mechanical details of "assessment forms".

14.Set the mechanical details of "when" the assessment results are given to the student.

15.Set the mechanical details of "who delivers the news".

16.Decide how much training, monitoring and feedback is given to the assessors.

All of these affect the design and development of the assessment instruments or activities. In addition, the creators of the assessment must address issues ofvalidity and reliability, frequency of testing, interpreting the results and special considerations.

5.2-1 Validity and reliability

Validity: is the result pertinent? does it measure what it claims to measure? does the score represent the student's ability? Validity has four dimensions.

Construct validity: is the basic hypothesis behind the test valid? (Example, a test designed to measure **changes** in problem solving skill with no explicit amount of excellent and relevant training.) Action: what's the hypothesis?

Content validity: does it test what you hoped? (Example, testing answers to a problem and claiming this tests problem solving skill.)

Concurrent validity: statistical comparison with other current valid test results for the same performance/knowledge/skill.

Predictive validity: statistical comparison between the current test results and future tests; does the test predict accurately the future performance? (Coefficients of validity are in the range 0.4 to 0.5.)

Reliability: is the result accurate? do we get consistent scores independent of the performance conditions or the evaluator and with sufficient sampling? Sometimes expressed as a correlation coefficient of reliability, **R**, that ranges from 0 to 1.0. A coefficient of 0.72 means that 28% of the variation in the score is attributed to error. We want a large number. Commercial tests have reliability >0.90. Your tests are OK if they are >0.60.

Test/retest (coefficient of stability): the same basic test to the same group but after elapsed time.

Two "equivalent" tests of the same group at about the same time and under the same conditions.

Internal test consistency: this is a measure to see whether all the questions on the test are of approximately the same challenge (split-half reliability). Spearman-Brown Prophecy test. Stone (1991) provides a good overview.

5.2-2 Test frequently and use variety

If all the testing is done at the "end" of the course, we call it summative assessment. Summative testing is frequently used for making a judgement. If we test throughout the course, we call it formative. Formative testing is used to guide activities. The trend, quite

correctly, is toward formative assessment. Plan to continuously or frequently assess your students.

When "competence" means a number of components (such as knowledge, interpersonal skills, technical skills, professional attitude and problem solving skills) it is unlikely that any **single** assessment method can be used. Decisions about student's progress require multiple assessments on multiple occasions using multiple methods.

5.2-3 Interpreting the results

Once we have the student's response do we compare it with our absolute criteria, select a "mark" and then see how everyone else in the class did? In other words, do we grade the results according to a normal distribution? The formal terminology is criteria-referenced versus norm-referenced. Criteria-referenced suggests we established measurable criteria for each goal and apply these to see the extent to which the goals have been reached. In one sense, these are "pass fail" decisions. In practice, we either set up the goals so that there is a range of goals or we measure the extent to which the criteria have been satisfied. Norm-referenced means we see how one student does relative to the distribution. In practice, we usually do a combination. If we set too difficult an exam so that all failed based on criteria-referenced, then we adjust the marks.

5.2-4 Special considerations: degree of explicitness on the task and on the process

Answers and the ease, reliability and validity of the assessment vary depending on the degree of "tightness" or explicitness built into both the task and the process.

Consider the **task**. The test question might ask "size a device to separate a mixture of...." or "size a cross-flow membrane to separate a mixture of..." The first statement is very loose; many options are possible. The second very specific or tight. If the task statement is loose, then a student might select a device that she knows a lot about but that may not be as pertinent as another. Another student might select the "best" device but be unable to size the device correctly.

Consider the **process**. The test question might say "for this separation problem, select the best device..." or "for this separation problem, list 20 optional devices and then write out three necessary and three desirable criteria. Use your criteria to select the best device." The first statement is very loose in terms of the process. We hope the person explicitly lists many options, generatesappropriate criteria and explicitly applies them. However, that's up to the person. In the second statement, the process to be used is spelled out; it is explicit. The process is tight.

Kimbell et al. (1991) report that providing a tight, well-structured set of process specifications dramatically improves the student's performance and tends **not** to discriminate well between poor and good skill. With tight specifications poor students are difficult to distinguish from "average" students. If the process is "loosely" described, then great divergence in marks occurs between poor and good students.

5.3 Some example "test" options.

Table 5-1 gives a general introduction to a variety of options, provides details and gives a qualitative rating of the validity and reliability. The ratings assume that you are working with tested instruments that are the best quality of the type. For example, poorly prepared multiple choice questions can have low validity and reliability. References are given that provide details.

In the rest of this Chapter, and in **Resources**, Section **F**, more concrete illustrations are given.

5.4 How do I assess subject knowledge?

Create objectives that have consistent, measurable criteria. Give the students many, varied opportunities to demonstrate their knowledge or skill. Example 5-1 illustrates the importance of explicit objectives.

Example 5-1:

Francine teaches a course on interest and depreciation. She asks if this is a good test question:

Qu.1:

"What is a current, reasonable value for interest on a high-risk loan?"

Answer:

I can't tell, Francine, until you spell out the course objectives. This fails criteria #2, 4, 5 and 7.

Comment:

Your question would be consistent with the following course objective: "Given a qualitative statement of risk, you should be able to estimate a reasonable value for the interest rate. Criteria: your answer will be within ± 2% of the answer developed by the instructor."

Table 5-1: Some options for assessment

Name	Description	Potential Validity	Potential Reliability	Context used	For more
Rating scale	Description of attribute being measured and	variable and often low.	variable and often low	attitudes, interpersonal skills	PED (1987)

	place for rating.				
Rating scale plus strengths & areas for improvement	Provides evidence for self assessment and improvement.				
Multiple choice questions (MCQ)	Consists of a "stem statement or question", plus a list of possible answers (called distracters) and the correct response.	high	high	subject knowledge and processing skills	PED (1987)
Word completion and short answer	Similar to MCQ without the prompts.	high	high		Swanson et al. (1991) PED (1987)
Oral exams	students orally respond to questions.	unreported	low but improve with structured oral exams plus standardizing the questions and defining clearly the competencies to be measured.	varies	Neufeld et al. (1985); Maatsch et al. (1984); PED (1987)
Direct observation	Observers view and critique student performance. Competencies being assessed determine the appropriate mechanism for observation.	requires more study.	high when performance situation is predetermined and clear "rating scales" have been devised specific for the competencies under scrutiny.	interpersonal skills and technical skills	Neufeld and Norman (1985); PED (1987); Grava-Gubbins et al. (1985); Van der Vleuten et al. (1985)
Essays	Students provide written	unsure.	improved by combining		Reiser (1980);

	description in response to a question.		major essay with a "learning contract", or keep essay brief with explicit and published criteria. Reliability improves if students are asked to write a dozen or so 10 minute essays on a variety of contexts.		Brown and Pendlebury (1992); Swanson et al. (1991)
Modified Essay Questions (MEQ)	Students are given an initial case and asked to respond to a progressive series of questions about a case problem. The information is given in a preset sequence. Example: given: brief problem; task: list hypotheses. Next task: list questions you would ask. Next, given the following new information, what is your hypothesis now. Thus, the sequence of questions and information revealed asks	reasonable validity.	relatively high reliability	subject knowledge and problem solving.	PED (1987)

	students to display reasoning and thinking processes as they are "led" through the problem solving process. No backtracking is allowed.				
Interrupted process: highlighters	Students shade over selected item; this reveals new information that allows them to proceed.				
Interrupted process: Triple Jump	Three, structured activities. 1) In response to a problem situation, students identify a prioritized set of issues for self study. 2) students self study and privately synthesize the new knowledge. 3) students respond to further questions to illustrate student's effectiveness. In the Medical School context, this is oral between a tutor and a student with 1/2 h on	untested.	acceptable. Swanson et al (1991) suggest that the use of only a few problems will not give reproducible and accurate measurements by the tutor of the problem solving, the self-directed learning or the knowledge.	process skills, self-directed and lifetime learning skills, subject knowledge.	Branda, Painvin et al. (1979); Powles et al. (1981); Henry et al. (undated); Swanson et al. (1991); PED (1987)

	#1; 2 h on #2 and 1/2 h on #3 which is a discussion with the tutor of the problem synthesis and conclusions student found from self study. Students self-assesses and the tutor provides feedback.				
Interrupted process: storyboard	Students are given a fold-over storyboard with timed tasks to do before they open a new section; tasks are written.	reliability and validity unreported but results improve if students write out their "reflections" as they progress.	process and subject knowledge	Kimbell et al. (1991)
Objective structured clinical examination (OSCE)	Like a practical MEQ: students rotate through a series of timed, 5 min "stations". At each station, they are given a different task to do: perform part of a physical examination, take a focused history, interpret patient data.	both reliability and validity for psychomotor skills low; relatively high on stations where the student writes a response. Examiners should compare standards and share ideas.	technical and psychomotor skills.	McKnight et al. (1988) Woodward et al. (1985); PED (1987)
Tutorial evaluation	Self, peers, and tutor assess at the end of each tutorial with on-going documentation.	questioned.	questioned.	varies.	PED (1987)
Learning contracts/	Written agreement between learner				Knowles (1964)

plans	and tutor as to objectives, learning strategies, timing, evidence to be produced, and assessment criteria. Learning contracts are not assessment tools, rather they are evidence that can be used in an assessment. They specify explicit and public agreement as to the expected assessment process.				
Written problem to solve	Students are given a series of problems to solve. The level of problem solving varies dramatically.			varies	
Portable patient problem pack (P4)	Students select from a series of cards the history, questions, actions, consultants; information is given on the back of each card selected. The card and the sequence are scored.	difficult to establish whether it is the subject knowledge or the processing skills.		varies	Barrows and Tamblyn (1980)
Simulations: written or		need 10 to 40 cases to			Swanson et al. (1991)

computer based		obtain reproducible results				
Simulations	Simulated patient: actress plays the part of a patient with a specific illness.					
Concept maps	Students complete a concept map or a Gowin Vee for the knowledge they are studying.					Novak and Gowin (1984)

Example 5-2:

For a course on interest and depreciation and objectives .... create questions that correspond with each of the first five levels of Bloom's taxonomy (Bloom's taxonomy is described in **HTGTM**, p. 7-7).

An Answer:

Bloom's level 1,2 definitions and comprehension:

Qu. 2 "Define interest."

Qu. 3 "Contrast Interest with depreciation."

Bloom's level 3 application

Qu. 4 "I borrowed $10,000 at 10% interest compounded daily. How much interest do I owe at the end of 30 days?"

Bloom's level 4 analysis

Qu. 5 "Attached is a solution to Question 4. Is it correct? If not, correct it."

Bloom's level 5 synthesis

Qu. 6 "Given the following course objective, create a reasonable, consistent 10-minute question for a test."

Comment:

Here we postpone creating the detailed objectives until we have created what we think are reasonable test questions. Then, we create the objectives and criteria. This is a reasonable approach to help create consistent assessment, goals, criteria and available resources.

Although questions at the higher levels require us to use "higher order thinking skills" (such as critical thinking and problem solving) the phrasing of the question does not allow us to see that thinking explicitly. We see the outcome or result. This is discussed in Section 5.5.

5.5 How do I assess process skills?

We use the same process that we used to assess subject knowledge: create objectives, measurable criteria and consistent, varied opportunities to demonstrate one's skill.

Some process objectives are given in **Resources**, Section **D.3**. Example questions are given in **Resources**, Section **F.1**.

Example 5-3

Michelle wants to test the student's ability to generate hypotheses. She suggests the following question:

Qu.7 List four issues which are both relevant to the attached problem and to the course learning objectives.

An Answer:

This question is consistent with the learning objective in "Lifetime learning skills:"

"Given a problem statement and the learning objectives in the course, you will be able to generate at least five learning issues that are judged to be 80% pertinent by the tutor."

However, Michelle should be addressing the learning objectives about hypothesis generation. She should **start** with the objectives. One of the learning objectives for hypothesis generation is:

"Given a problem situation, as an individual you will be able to write out 50 ideas in 10 minutes; these will contain at least seven different classes of ideas and at least two classes of ideas will be judged by the tutor to be pertinent."

Hence, Michelle should be clearer in her mind as to what she wants to test. **Qu. 7** asks the student to provide an outcome. Such an outcome is vital for the process of Lifetime learning. **Qu. 7**, however, does not give the assessor a chance to see the process used by the student in obtaining the outcome. If the hypothesis generation process is being assessed, then she might reword her question:

Qu.8 List four issues which are relevant to both the attached problem and to the course learning objectives. Display the process used by:

a. listing all your ideas (10 minutes);

b. classify your ideas and indicate the basis of your classification;

c. list your criteria for selecting the relevance to both the attached problem and the course learning objectives;

d. write out your top four issues.

Comment:

For the objective listed, only part **a** of the question is pertinent. The task of listing issues is complex involving brainstorming, criteria generation, classification and decision-making.

5.6 How does assessment link to objectives?

Assessment must link directly with the objectives. Without objectives, you cannot assess fairly. As discussed in Section 4.5, this poses a problem for programs where students have the freedom to create their own objectives. When students generate their own learning objectives, then

we have to have access to those objectives if we are doing the assessment, or

we empower the students to do the assessment and we monitor the process.

5.7 How does assessment link to resources?

Assessment is linked to objectives. The objectives must be realistic. The resources must be available for the students to be able to achieve the objectives.

5.8 On the role of evidence

Example 5-4

Sara completed a feedback form about her contribution to the group. She assessed herself as a 9. The tutor completed the same feedback form and assessed Sara's contribution as a 6. The form included observable objectives and measurable criteria. Both were surprised and frustrated when the tutor sat down with Sara to go over her performance.

Comment:

Both are assessing on visual impressions of what occurred in the group and recording their impressions on the same form. Consider how different the process would be if, three times during the meeting, all participants took a 3 minute time-out while each wrote a reflection on what had been going on in the group, where each thought the group was headed, and what each thought his/her contribution had been. Further, suppose that at the end of the session, each group member completed a feedback form for every other member. Now suppose Sara was then asked to complete her own feedback form using all the above data as evidence.

In this case, the tutor can consider Sara's assessment in the context of evidence.

Example 5-4 helps us to appreciate the role of **evidence** in doing any assessment. Evidence gives the basis for the assessment. True, the goals and measurable criteria must be clear. However, we also need a variety of evidence.

As we consider the types of evidence we can help the student gather, we can modify our goals and criteria so that these three form a consistent package.

One of the simplest and most helpful forms of evidence is the single page reflection. Some mistakenly believe that the interrupting the process of learning to reflect is too disruptive. Contrarily, Kimbell et al. (1991) have shown that interrupting the design or problem solving process to reflect actually improves one's problem solving and thinking processes. Indeed, my recommendation is to ask students to pause and reflect often as they work through a case. Try to arrange the reflection activity when the students have completed a phase of the process -- instead of arbitrarily asking them to reflect every 45 min of process time. Reflection is linked to monitoring. Since monitoring occurs near the beginning and near the end of a task, use these times to ask the students to write down their reflections.

5.9 Peer assessment? How to get students to assess each other instead of patting each other on the back? or putting all the others down so that they appear to be good?

Peers are a valuable source of feedback. Peer feedback can be prompt and individualized. Given that few have much experience with peer feedback:

We need to train students so that they understand the assessment process. The key ideas are: goals and criteria must be explicit and measurable; clients being assessed need to have an opportunity to gather/present evidence that is pertinent to progress toward the goal; the assessment is a judgement of the pertinence, quality and extent to which theevidence substantiates any claims to be made about achievement.

We also need to provide an environment where peers give accurate feedback. Some tend to rate colleagues - especially friends - **high** because they think they are doing them a favour. In reality they are **not** helping but are misguiding through misinformation. Some tend to rate others **low** so that they, by comparison, will look high.

Training

Table 3-1 describes a workshop to improve assessment skills (in the MPS unit **3**). More about principles upon which assessment is based are given in **HTGTM** Chapters 8 and 9.

I suggest that students and tutors experience a 3-6 h workshop on assessment.

Environment

Create an environment that rewards fair and accurate assessment. We can do this by making assessment a learning objective for the course. That is, students are assessed on their ability to assess others. In other words, students assess peers and their assessments are marked by the tutor. Gradually, as students develop consistency and objectivity, their assessments can be given directly to their peers.

In other words, the peer assessment is never considered to be "let's say nice things to our friends." Rather it is a skill under development.

Contrary to my belief in the merit in trained peer assessment, Swanson, Case and van der Vleuten (1991) suggest that peer and tutor assessment of the process of learning are not appropriate. I suggest that the difference is in the word "trained."

Stefani (1994) found that in an environment of a clearly defined, carefully monitored assignment, the student's self-assessment was at maximum 3% below the tutor's mark. The self-assessment was always lower than the tutor's mark over the whole range of student abilities. For peer assessment, the marks were about 3% lowerthan the tutor's marks for students with marks in the range 45% to 62%; the peer marks were about 2% higher than the tutors marks for all students with marks greater than 63%.

Section **D.3** of the companion book **Resources** lists learning objectives for acquiring skill in self-assessment as is used in MPS unit **3**.

5.10 Self-assessment?

Example 5-5: The tutor, Jean, has previously reviewed Ahmed's reflective journal for this unit on creativity. She sees the evidence as supporting a claim of being able to do "most of these" or a mark of 8/10. Jean now has a personal interview with Ahmed.

Before the interview, Ahmed has collected his evidence for his accomplishments for developing his creativity. His conclusion is that he has achieved **all** of the goals for MPS Unit **7**. This would give him a "mark" of 10/10. In the interview, Ahmed presents this evidence and adds, "I see myself as 10/10." The evidence that Ahmed presents is the same as Jean has seen. No new evidence is given. How does Jean handle the rest of the interview?

An Answer:

Jean: "Ahmed, I am delighted that you see yourself as achieving **all** of the objectives. I saw, in your reflective journal, evidence to support "most of these." In particular, for Objective 3.2 I found evidence that you could think of 35 ideas in 5 minutes, whereas the course objective was 50 in 5 min. In addition, I found evidence that you had could ... (here Jean refers to specific objectives and their criteria). What am I missing? I need to see evidence to support claims."

Comment: Jean's response depends on Ahmed. In my experience, 80% of the students accept Jean's focus on evidence, and the resulting assessment. The other 20% of the students may focus rather on the **mark** of 8 instead of 10 and argue that they have done better. "They deserve the 10."

Here are some options Jean has:

a) reconvene another interview after Ahmed has had a chance to gather appropriate evidence.

b) give Ahmed a chance to provide evidence in the interview by giving him a problem and asking him to demonstrate his skill.

c) accept his "mark" on creativity and use this interview as evidence for his inability to self-assess. His mark on creativity might be 10/10 but his mark on self-assessment is 5/10.

****

(Background details about the objectives for MPS Units **7**, on creativity, and Unit **3**, on self-assessment, are given in the companion book **Resources**, Section **D.3**.)

The same issues and approaches used for peer assessment apply to self-assessment. Provide training, give feedback and focus the tutor's assessment on "the quality of the evidence and how well it substantiates the claims of a student.

Example 5-5 illustrates an interview with the student.

5.11 Can tutors accurately assess student performance in groups of five to ten?

No. Swanson, Case and van der Vleuten (1991) suggest that tutor ratings carry little measurement information and have poor validity. I think that can be corrected if the shift is on an assessment of written evidence and tutors are trained in assessment. The key stumbling block is the **number** of performances we are expected to assess simultaneously. Researchers in group process suggest that one observer cannot accurately assess more than two people simultaneously (de Stephen, 1985).

So what might we do? My approach is to train tutors, and students alike in the assessment process; make high quality assessment and self-assessment intended outcomes of the

program. Provide numerous opportunities for peer and self assessment so that each individual regularly collects evidence and includes this in a reflective journal. The tutor's role is to assess the student's claims based on the evidence.

Other approaches include the use of Feedback Meetings and Elaboration Meetings (described in Chapter 4, Table 4-2).

5.12 Why assess the process skills?

Why assess? Because process skills are required in small group, self-directed, interdependent PBL, andbecause if we don't make these process skills valuedoutcomes, little skill will be acquired.

5.13 Assessing large classes

Two options are available for assessing large numbers of students:

The tutor does it but structures the assessment and feedback so that it takes between 3 to 10 min to assess each assignment. Example 5-6 illustrates an approach.

Example 5-6

About 60 students complete 10 to 15 page reflective journals for each MPS Unit. This means at least one set of reports per week. With three tutors, we each have 20 reports to grade per week.

An Answer:

We created a standard format for the reports; a convenient feedback checklist form and could provide valid and accurate feedback spending fewer than 10 min per report. Initially, we double marked about 5 reports to ensure that we had inter-rater reliability.

The tutor can empower the students with the task of assessing and giving feedback. The tutor's task is now to make the process visible so that the students are accountable (their work can be monitored). This Chapter emphasizes this option.

5.14 How do you integrate peer and self-assessment into the tutor's mark?

The answer depends on how much emphasis you place on peer and self-assessment as outcome skills. If high quality peer and self-assessment are **not** objectives, then students are free to give high marks to their friends or give low marks so that their own will appear high. The temptation is for them to manipulate the system.

If the students have been trained and self-assessment is an expected outcome (with all the empowerment and accountability described in this Chapter) then I have had very good success with the following approach:

peer-**assessment** does not contribute directly to the overall mark but becomes evidence for high qualityself-assessment.

Table 5-2 Example assessment

All required assignments are due before class on the due date. Assignments handed in late receive zero marks.

All assignments must have a covering letter or memo that describes what the task was, what answer you obtained and some judgement as to the accuracy and whether the reader should put much credence to your answer. For any assignment, I may read and evaluate the complete assignment based only on the brief, one page memo or covering letter.

Wherever possible, the assignments are evaluated in terms of what would be expected in professional practice. This applies to the format and clarity used in presenting your results, in the judgement you exercise in commenting on the practicality and reasonableness of your answer.

The final assessment in this course is the highest combination of a 40%- 60% weighting for term work and for the final exam. You may elect to have any other weighting between 10% term work to 90% term work. Your personal choice should be made in writing to me before Nov 1.

The final exam is an open book exam.

The term work is 30% self-assessment. The self-assessment is based on the evidence obtained from peer and self-reports written in workshops. The evidence is included in your reflective journals that you submit throughout the term and confirmed in a private interview with the tutor at the end of the term. Because much of what is learned is through participation in class or tutorial, you are expected to attend and participate in **all** classes.

For the Unit on self-directed PBL, failure to attend and do your share of the activities for all of the sessions will result in a failure in that Unit on self-directed learning. You are required to learn the subject knowledge being addressed in those PBL activities.

the **self-assessment** includes the feedback from thepeers as part of the evidence for a reflective journal writeup. The journals are marked for the objectivity and completeness in interpreting the evidence.

Once the quality of the journals was high, we could use these with confidence to assess the degree to which the process skills had been achieved. In addition, the students prepared sample examination questions including their solution; the students attempted to solve questions posed peers. Peers marked the results. This "feedback" component was monitored by the tutor. In this context, the self-assessment component contributed 30% toward the term assessment; the tutor's mark contributed 70%. The course also had a

summative written examination. The students could "contract" (no later than midway through the term) to have the term component count any percentage between 10 and 90% toward the final contract grade. The default contract was the best of a 60-40 split. Table 5-2 describes an example approach to assessment.

Therefore, in my courses I have four different marks:

1. Exam mark. This summative mark is from the final, 3 hour written examination.

2. Self-assessment mark. This summative mark results from the student-tutor interview. This mark contributes 30% toward the term mark.

3. Term mark. This is 30% self-assessment and 70% the marks of the student journals written throughout the year.

4. The contract mark. This depends on the individual contract between the tutor and the student. The contract mark can be any combination (in the range 10 to 90%) of exam and term marks. Thus, a student's contract mark might be 15% exam plus 85% term marks. If the student elects not to formally contract, then he/she receives the highest combination of a 40 to 60% weighting for the exam and term marks.

Table 5-3 shows, for two courses, the average (and [standard deviation]) for the exam marks, the contract mark (based on whatever weighting for term mark that the individual student selected), an example contract mark based on a 60% weighting for term mark and another example based on a 40% weighting for term mark.

Table 5-3 Some typical results including self-assessment, contract for the weighting for formative assessment versus summative written exam

N= 49 with 75% electing contract >70%; 0%<40% formative

Exam marks75.85 [14.90]

Contract mark 81.73 [5.99]

60% formative78.73 [9.66]

40% formative77.77 [8.76]

Contract-self 1.09 [5.51]

Contract-exam 5.88[12.65]

N=50 with 15% electing contract >70%; 0%<40% formative

Exam marks79.26[11.57]

Contract mark78.2[10.5]

60% formative77.0[11.3]

40% formative77.8[11.0]

Contract-self 0.15 [5.3]

Contract-exam-1.13 [5.17]

For the first course, the contract mark was, on average, 1.09 marks higher than the self assessment with astandard deviation of 5.51. For the second course, the contract mark was only 0.15 marks higher than the self-assessment mark.

Table 5-4 Comparison of peer and self-assessment marks with teacher's marks.

(Brown and Pendlebury, 1992)

within under over

± 10% estimate estimate

-10% +10%

_____ _____ _____

Self 76% 23% 1%

Peer 80% 6% 14%

Another way of looking at the marks is to note the percentage of students whose self assessment marks are within a certain percentage of the "final" mark. Table 5-4 summarizes the work of Brown and Pendlebury (1992) that shows that the self-assessment mark is within ± 10% of the "final" mark; Table 5-5 shows that 78% of the self-assessment marks are within ± 5% ofthe final mark.

Table 5-5 Comparison of self-assessment mark with "contract mark" N=55

within under over

± 5% estimate estimate

-10% +10%

_____ _____ _____

Self 78% 2% 0%

In summary, we can use a variety of methods of integrating the peer, self and tutor's marks. I interpret the results in Tables 5-3, 5-4 and 5-5 to mean that there is little difference between the self-assessment and the final mark; indeed the self-assessment tends to be slightly less than the final mark. I believe that empowering students with a significant amount of control over the assessment components improves their performance.

5.15 Pass/fail? or Marks?

For those of us used to assigning marks, it comes as a surprise that one might even consider replacing the mark with a "pass/fail" decision. However, when students are given marks in a PBL course (and to a lesser extent in other competitive learning environments) the following events have occurred,

- books are stolen from the library,

- student assignments are stolen (both from the "in tray" - so that it will appear as though that student has not done the work and from the "marked" tray- so that others will benefit from seeing the results of good assignments).

- chapters are torn from books,

- articles are torn from journals.

In other words, the competitiveness associated with marks has translated into underhanded tactics of hoarding resources.

Hence, many medical and pharmacy schools have gone to a Pass/fail systems. Under this system librariansreport little evidence of resources being defaced or hoarded.

Pass/fail seems to be an assessment tactic to try to promote cooperative learning rather than intense competitiveness.

Approaches toward **marks** are

1. convert the course to Pass/fail.

2. give 100% for completion of task. (In other words, give full marks for participation.)

3. convert "marks" into other terms:

T (meaning <50%) "the tutor is in trouble because you are not learning as much as expected;"

Y (meaning 50 to 60) "you can do better."

K (meaning 61 to 79) "OK."

Z (meaning >80) "Terrific."

4. change marks to feedback on five "Strengths," two "Areas to work on" and three "Insights."

The students are expected to maintain their strengths and shift one of their areas to work on into a strength.

5. give marks to the team effort; all members of the team get the same mark. Devise methods for sampling performance by asking individual members of the team to display skill. This tends to minimize the competition within teams.

6. make self-assessment a skill to be learned in the course and assess the student's ability to self-assess.

Hoarding/defacing the resources

Here are some options:

1. have enough resources available that all groups have a copy of the key resources.

2. library reserve policy with tracing for defacing.

3. use the jigsaw approach or a Consolidate meeting, described in Sections 4.2 and 4.11, where those studying the same topic get together to become expert. Use assessment techniques for this activity that promote cooperation.

4. change student attitude toward assessment so that positive assessment is **not** achieved by putting othersdown but rather by improving self. Make explicit the implicit and focus on assessing the student's ability to assess. Empower the student with the task of assigning marks to themselves and to their peers while the tutor assesses the process used.

Approaches to assessment and training students in the assessment process are described in **HTGTM** p. 8-4 and 8-5.

5. let the students personally elect the assessment process they wish. This was described in Tables 5-2 and 5-3.

5.16 Summary:

Assessment is important because students learn what is assessed. In any assessment, the issues of interest are the purpose and mechanics of the assessment, the validity and reliability of the test, the variety and frequency of the testing, special considerations and methods for interpreting the results. About 20 test options are tabulated.

The rest of the Chapter considered a series of questions relating to how to assess subject knowledge, process skills, the role of evidence and peer and self-assessment.

5.17 References

Bloom, B.S. (1956) "Taxonomy of Educational Objectives: classification of educational goals: Handbook I: Cognitive Domain," McKay Publishers, New York, NY.

Boud, David, ed (1981) "Developing Student Autonomy in Learning" Kogan Page, London

Boud, David (1981) "The Role of Self-Assessment" University Education News, vol 2, No. 2, p 10 December

Boud, David (1986) "Implementing Student Self-Assessment" Green Guide No. 5, HERDSA, Tertiary Education Research Centre, the University of New South Wales, Kensington, NSW Australia

Branda, Luis (1992) Biochemistry, McMaster University, Hamilton, personal communication.

Brown, G. and M. Pendlebury (1992) "Assessing Active Learning," Parts 1 and 2, CVCP Universities' Staff Development and Training Unit, University House, Sheffield, S10 2TN, UK.

de Stephen, R.S. (1985) Sinclair Community College, personal communication.

Edel, R.L. (1979) "Essentials of Educational Measurement," 3rd ed., Prentice-Hall, Englewood Cliffs, NJ.

Feletti, G.I. (1980) "Reliability and Validity studies on Modified Essay Questions," J. Medical Education, **55**, 933-941.

Feletti, G.I. and C.E. Engel (1980) "The Modified Essay Question for Testing Problem-solving Skills," Medical J. Australia, **1** 79-80.

Ferrier, B., M. Marrin, and J. Seidman (1981)"Student Autonomy in Learning Medicine: Some Participant's Experiences" in "Developing Student Autonomy" D. Boud, ed., Kogan Page, London.

Grava-Gubbins, I., S.B. Khan and P. Raspberry (1985) "A Study of the Reliability of the 1985 Simulated Office Orals," The College of Family Physicians of Canada, Toronto, Oct.

Harden, R.M., S. Sowden and W.R. Dunn (undated) "Some Educational Strategies in Curriculum Development: the SPICES model", Centre for Medical Education, University of Dundee, Dundee, Scotland.

Henry, N., R. McAuley, G.R. Norman, V.R. Neufeld, P. Powles and P. Dodd, (undated) "The Triple Jump Exercise- reliability and validity, McMaster University Medical School, Hamilton ON internal report.

Kimbell, R., K. Stables, T. Wheeler, A. Wosniak and V. Kelly (1991) "Assessment of Performance in Design and Technology," School Examinations and Assessment Council, London W11 3JB

Knox, J.D.E. (1980) "How to use Modified Essay Questions," Medical Teacher, **2**, 20-24.

Maatsch, J.L., R.L. Huang and S. Downing (1984) "The Predictive Validity of Test Formats and a Psychometric Theory of Clinical Competence," Proc. 23rd Conference on Research in Medical Education, Nov., Washington, DC

McKnight, J., E. Rideout, B. Brown, D. Ciliska, D. Patton, J. Rankin and C.A. Woodward (1987) "The Objective Structured Clinical Examination: an alternative approach to assessing student clinical performance," J. Nursing Education, **26**, 1, 39-41.

Neufeld, V.R. and H. S. Barrows (1974) "The McMaster Philosophy:

an approach to medical education", EducationMonograph No.5, Faculty of Medicine, McMaster University, Hamilton, Ontario.

Neufeld, V.R., and G.R. Norman (1985) "Assessing Clinical Competence," Springer, New York, NY.

Neufeld, V. and Sweeney, George (1984) "M.D. Program General Guide 1984" Faculty of Health Sciences, McMaster University, Hamilton.

Novak's "Helping Students Learn How to Learn: a view from a Teacher-researcher" (1989) (reviewed PS News **69**) suggests that we can ask Methods of providing assessment and feedback about a concept map are given in Chapt 5 of J.D. Novak and D.B. Gowin's "Learning How to Learn" Cambridge University Press, 1984; and by Karl Smith and colleagues "Constructing Knowledge bases" (1985) reviewed in PS News **73** p 17. Reliability and validity are not reported. Competencies: knowledge structure.

Painvin, C., V.R. Neufeld, G.R. Norman and L. Walker (1979) "The Triple Jump Exercise- a structured measure of problem solving and self-directed learning," Proc. 18th Conference on Research in Medical Education, Nov., Washington, DC

PED (1972) "Evaluation of Student Progress" MD Programme, Faculty of Medicine, McMaster University, Hamilton, ON Education Monograph No. 1, March 1972.

PED (1987) "Evaluation Methods: a Resource Manual", Program for Educational Development, Faculty of Health Sciences, McMaster University, Hamilton, (1987) is an excellent review of options

PED (undated) "The Small Group Tutorial" M.D. Program, McMaster University Hamilton.

Powles, A.C.P., N. Wintrip, V.R. Neufeld, J.G. Wakefield, G. Coates and J. Burrows (1981) "The Triple Jump Exercise- further studies of an evaluative technique," Proc. 20th Conference on Research in Medical Education, Nov., Washington, DC.

Reiser, R.A. (1980) "An Objective Approach to Scoring Essays," J, of Instructional Development, **3** (4) 19-22.

Rothman, D.S., D.E. Tobias and S.M. Speedy (undated) "Validation of Written Simulations as Measures ofProblem Solving for Pharmacy Students," School of Pharmacy University of Maryland, Baltimore MD 21201

Stafford, P. and H. Pierce-Fenn (1985) "Predictive Validity of the Modified Essay Question," Physiotherapy Canada **37**, 6, 356-359.

Stefani, L.J. (1994) "Peer, Self and Tutor Assessment: relative reliabilities," Studies in Higher Education, **19**, 1, 69-75.

Stone, H.L. (1991) "From Analysis to appraisal: measurement skills for self-management," in "Data Gathering Instruments for Evaluating Educational Programs and Teaching Effectiveness in the Center for Health Sciences, University of Wisconsin (Madison), Madison, WI.

Streiner, D.L. (1985) "Global Rating Scales," in "Assessing Clinical Competence," V.R. Neufeld and G.R. Norman, eds., Springer, NY., 119-141.

Swanson, D.B., S.M. Case and C.P.M. van der Vleuten (1991) "Strategies for Student Assessment," Chapter 27 in "The Challenge of Problem-based Learning," D.Boud and G. Feletti, eds., Kogan Page, London.

van der Vleuten, P.M. and S.J. Van Luyk (1985) "A Validity Study of a Test for Clinical and Technical Medical Skills" Proceedings of the International Conference on Newer

Developments in Assessing Clinical Competence, World Federation of Medical Education. Ottawa pp 117-133.

Woods, D.R., R.R. Marshall and A.N. Hrymak (1988) "Self-assessment in the Context of the McMaster Problem Solving Program," Assessment and Evaluation in Higher Education, **13**, 2, 107-127.

Woodward, C.A., D. Patton and B. Brown (1985) "The Relationship Between Nursing Student's Evaluation of the OSCE and Performance," Proceedings of the International Conference on Newer Developments in Assessing Clinical Competence, World Federation of Medical Education, Ottawa.

Back to PBL