

Assessment in Problem-based Learning (PBL)

S J van Luijk,^{*MD, PhD}, C P M van der Vleuten,^{**PhD}

Abstract

Problem-based learning (PBL), a relatively new approach to teaching, has been introduced in several of the medical schools in the Southeast Pacific region. This approach emphasises active, self-directed learning and puts the emphasis on the individual student and his/her role in the educational process. It is well known that assessment plays a large role in influencing student learning behaviour. Therefore, it is important that the assessment process not hamper learning or adversely affect attainment of the goals of the curriculum. If student behaviours are directed toward achieving success on the assessments, instructors' efforts to create a climate of self-directed learning and individual responsibility will be frustrated. This article provides information on the different classes of assessment vehicles available to assess the various domains of medical competence. We pay special attention to the critical relationship between the educational goals of problem-based learning and the different assessment instruments used. We briefly describe some assessment instruments and how it can relate to these goals in a positive manner. We conclude that no assessment instrument is ideal and that each has unique inherent deficiencies. As a consequence, a successful student assessment process requires use of several complementary instruments.

Ann Acad Med Singapore 2001; 30:347-52

Key words: Clinical competence, Metacognitive skills, Problem-solving

Introduction

Assessment of clinical competence has always received considerable attention in medical education. Teachers and students frequently challenge the value of assessment and are always looking for alternative methods of assessment. Since it is imperative that the assessment systems reflect the educational goals, debates about assessment were even more prominent when problem-based learning (PBL) was introduced as an alternative to more traditional ways of teaching.

These considerations made apparent the need for educational research on assessment in higher education. A wealth of research has indeed been conducted during the last few decades and now we know much more about assessment and its role in curriculum design.

PBL is characterised by learner-centred education and so it is important that the assessment process be tailored to fit the learning process.

PBL also emphasises integrated learning activities. Integration can be achieved between disciplines, between knowledge and skills and between knowledge, skills and

professional behaviour. This means that in problem-based programme, assessing higher cognitive abilities and professional performance is very important. The challenge, therefore, is to ensure a fair match between the educational goals and the driving force of assessment.

This communication begins with a description of the domain of assessment of clinical competence. Next, different classes of assessment instruments are matched to different components of clinical competence. On the basis of these descriptions, suggestions for assessment in PBL are given.

The Different Aspects of Clinical Competence

For many years competence has been seen intuitively as a compilation of different and more or less independent components, including knowledge, problem solving, clinical skills and attitudes. These components have been more or less regarded as separate "traits" of a person. This view has led to the development of assessment instruments for each of the different components. Any overlap between the components has been seen as a negative feature of the instrument. Success in each component is said to reflect the level of competence of the student. Critics

* Assistant Professor

** Professor of Education and

Chair Department of Educational Development and Research

Medical School, Maastricht University, The Netherlands

Address for Reprints: Dr S J van Luijk, P O Box 616, 6200 MD Maastricht, The Netherlands.

E-mail: S.vanluijk@educ.unimaas.nl

of these instruments indicate concerns about their validity.

If the separate components are independent entities, one should expect high correlations of results within the components and low correlations of results between the components. However, correlations of scores from measurement instruments within and between components of competence do not meet these expectations.¹ No specific correlation patterns are observed.

Miller² has suggested another model of clinical competence. He describes clinical competence as four layers of a pyramid. Each layer is tightly adherent to the one below and the one above. The first layer refers to the knowledge needed to carry out professional tasks. This knowledge comprises facts and its assessment is largely covered by a range of so called “objective tests” like multiple choice.

The second layer is called the “knows how” and refers to the application of knowledge in context, for example in a certain problem or clinical situation. The third layer is the “shows how” level. At this level the student must show that he can perform in a simulated or artificial practice setting. The top of the pyramid or the highest level of competence is the “does” level, representing the student in action in medical practice.

We will discuss the measurement issues in each level of Miller’s pyramid, some outcomes of relevant research and some illustrations of measurement methods used to assess the competencies at these levels.

Assessing the Lower Levels of Competence: The “knows”

As stated earlier the “knows” refers to factual knowledge. This contrasts with the second layer in which the application of knowledge to a problem is important. Many methods are used to assess factual knowledge including multiple choice questions (MCQ), oral exams or essays.

Frequently, a knowledge test is used as an end-of-course test. During the course, students will focus on those learning goals and study activities needed to pass the test. In PBL, there is a tension between the end-of-course assessment with curriculum-dictated fixed objectives and the responsibility of students to formulate their own learning goals.

Although experts have devised no single solution to this conflict, to a certain extent the problem can be managed. The first approach is to ensure that not every test has consequences for passing or graduation. Test results can be used solely to inform students about their level of knowledge (formative assessment). In line with the tenets of PBL, students can then take the responsibility to correct their weaknesses. Formative assessment will clearly diminish pressure on students in preparing for the test, therefore

giving more “space” for the learning process in the tutorial group.

Another alternative deals with the development of knowledge tests that have little impact on student learning behaviour. An example is the progress test used at the University of Maastricht. The progress test is a test that covers the whole knowledge domain of knowledge (all disciplines) and is targeted at a level of competence expected at graduation.³ It is a multiple-choice test with true/false questions and a “don’t know” option. It contains questions from all clinical disciplines, basic disciplines and relevant disciplines of other health sciences. Every student of every class is expected to take part in the test. Of course, it is expected that students in the first year will perform at a lower level than their senior colleagues so they are allowed lower pass/fail scores. The test is taken four times a year. Students’ scores are determined by the true minus false score. Guessing is penalised with a false answer counting as minus one point. A correct answer is plus one point. Because students’ knowledge increases during the curriculum, the scores of the test constitute together a “growth curve”.

Since the test covers the whole content of medical knowledge, students are discouraged from studying for the test, because they cannot expect to master the content by learning it in a few days or even a few weeks. In contrast, learning regularly for the tasks of a tutorial group is the best strategy to produce satisfactory test results. Furthermore, extra learning activities will be rewarded because more knowledge will lead to successfully answering more questions and therefore provide a greater chance for higher scores.

Since all disciplines are represented in the test, it is counterproductive for students to neglect essential parts or disciplines of the curriculum. They must continue to have a firm grasp on the knowledge of pre-clinical years like biomedical sciences or anatomy right up to graduation. For these reasons we feel strongly that progress testing is a very appropriate way to assess knowledge in a problem-based curriculum.

Assessing the Application of Knowledge: The “knows how”

Problem-based learning teaches students to solve problems within a meaningful context. Therefore, applying the knowledge to solve clinical problems (one step higher in Miller’s pyramid) is an important issue in assessment of PBL.

To assess clinical reasoning ability in the past, attempts have been made to simulate real patient encounters using written simulations. These simulations attempted to reflect the several steps in the decision-making process of a

physician dealing with a patient. In these situations, the student was required to collect data on history-taking, physical examination and laboratory data. The test taker's problem solving ability was judged by the degree to which he or she followed the decision-making process an expert would use.

Partly because written simulations were realistic, they became very popular. Examples of instruments are: patient management problems.⁴⁻⁷ Despite the intended enthusiasm for those instruments, several drawbacks become apparent following their implementation.

The first problem was that student performance was difficult to score. For example, questions arose as to how to score students who followed a decision-making process different from that of the experts, how to weight wrong answers, how to correct students when they made mistakes during the process? Compounding this was the difficulty getting consensus among experts about the right way to handle a case.

The second problem was that student performance on one case was not predictive of performance on another case even within the same domain. Problem-solving skills were closely related to the content of the case. A generic trait such as problem-solving capacity did not exist. This has been called the problem of "content specificity".⁸ These findings have important implications for the assessment of problem solving. Many cases (patient problems) are needed to achieve an acceptable reliability, that is, a score that is independent of the sample of cases used in the test. To simulate total patient encounters, a considerable amount of testing time is needed for each case. Therefore, to get reliable data, a number of cases are needed and this requires very long testing times. Because of the fact that resources are usually limited, problem-solving tests are often too short, test too few cases and are therefore unreliable.

The third problem stems from the fact that experts do not perform better on simulations than students.⁹ This raises serious questions about validity.

The last unexpected finding was that if problem solving was measured reliably, using many cases, the results correlate highly with other assessment instruments including multiple-choice tests.¹⁰

All of this has important implications for practice. To counter the content specificity problem, more efficient assessment methods were required. In most patient encounters, one or two decisions are essential. The remaining part of the interaction is either less relevant or follows from the essential decisions. It may, therefore, be appropriate to focus on the essential parts and ignore the rest. This has been called the "key-feature approach".^{11,12} This more efficient assessment method might be expected to reduce the testing time per case and, therefore, for

the examination as a whole. This means that more cases can be offered in the same testing time, which enhances reliability.

Another alternative is to increase the frequency of testing and combine the results. In this approach, the content is tested over weeks or months. The content is spread out over all the measurement moments. Regrettably, such a procedure is still time-consuming and costly.

Research tells us that what is measured is much more important than how it is measured. The content of the question or task (the stimulus) and not the format of the question determines what cognitive activities students perform.¹³ Multiple choice questions, essays or orals can all assess problem-solving abilities of students depending on the content of the question.¹⁴ The content should reflect practice so it has to be "contextually rich". An authentic context gives students clues to recall the information they have stored in long-term memory.

In PBL, authentic situations (problems) are offered to students in an effort to improve integrated learning. Learning with problems is critical to the success of PBL so obviously assessment of problems has a high priority. Bearing in mind the research findings, the current approach to student assessment is to focus on simple, short, clearly described authentic cases. Context is provided in the case without interpretation and students deduce problems directly from the content of the case. The content of the question determines the format.¹⁵ The application of modern multimedia enhances the authenticity and facilitates assessment in different locations (i.e. different teaching hospitals).

Assessing Professional Skills: "shows how"

Showing what students have learned in simulated practice is the third of Miller's layers. This is an important domain for assessment in PBL since it looks at the integration of knowledge and skills into a real life simulated situation. Observing actual behaviour is the important extension compared to the previous layers.

Different assessment instruments are used to observe students in simulated practice. Examples include Objective Structured Clinical Examinations (OSCE);¹⁶ and standardised patient-based tests.¹⁷

The typical assessment is to put students in simulated clinical situations and observe their performance. Examples of skills to be assessed include physical examination skills, laboratory skills, emergency skills and communication skills. In a performance-based test, students enter a room, or station, in which they have to perform a skill. Depending on the content of the station, a teaching model or a simulated patient is present in the room. A simulated patient is trained to play the role of the patient as authentically as possible. While completing the task, the student is rated by an

examiner or by the simulated patient. The rating is done with a predetermined rating scale or checklist. After some time, varying mostly from 5 to 20 minutes, the student has to enter another station to perform another assignment. Depending on the test length, the student may rotate through a number of such stations.

Assessing the “shows how” in principle is not very different from assessing the “knows how”. In the “knows how”, confronting students with authentic contexts assesses cognitive abilities. The same is done in the “shows how” but now on a behavioural level. Because both types of assessment are closely related, the advantages and difficulties with both are quite similar.

Students put a considerable amount of energy into practising the “shows how” skills¹⁸ and time spent practising increases dramatically before the test.

One of the difficulties in the “shows how” assessment is the content specificity problem, which shows up in this layer as it did in the first two. Performance on one station does not predict performance on other stations. Therefore, many stations are needed to achieve acceptable reliability.

Scoring problems also exist. With performance-based tests, examiners are inclined to assess students with detailed checklists of separate skills in the hope of limiting differences between examiners. Students have responded in different ways. Prior to taking a test, students often study the checklists given them by friends who have previously undergone a similar assessment. This manner of preparation may have negative consequences. Firstly, “studying to the test” leads to superficial learning of information. It is not committed to long-term memory and thus is easily forgotten. Secondly, it leads to aberrant performance wherein the student carries out the requested task in a stereotypical manner virtually disregarding the patient and the patient’s physical findings. Finally, running such performance tests is quite costly compared to written tests, although the average school is still capable of managing the costs.^{19,20}

Just as the problems in layer three are similar to the previous layers, so are the solutions. Overcoming the ubiquitous problem of content specificity requires examiners to: assess the key elements of the content of the case; avoid constructing complex cases; strive for authenticity.²¹ Learning checklists by heart can be countered by including more global assessment items on the checklist. The consequence of this is that feedback will become less specific. However, assessing students with the use of more global items has no severe consequences for the total test reliability.²²

In PBL, performance-based tests have become very popular because the content is directly related to actual

performance. As previously mentioned, improvement of these tests can be achieved by including simple problems, using more contexts in the problems and focussing on key decisions. Because the assessment stimulates practising skills, it fits very well within a problem-based environment.

Assessing Performance in Authentic Situations: The “does”

Because PBL prepares students from the start of their studies for performance in real life situations, the assessment of performance in real life deserves attention. In most cases, this assessment takes place in the context of clinical rotations in the later part of the medical school curriculum. In contrast to what occurs in the “shows how” layer, the “does” layer focuses on an ongoing observation of student performance in authentic situations. Regrettably, these kinds of assessment are not appropriately applied.²³ Three serious limitations affect the assessment of “does”.² Firstly, there is a lack of standardisation in assessment. In general, no predetermined criteria exist for assessing directly observed students. Moreover, most supervising physicians are not well trained to observe and assess students. Secondly, there is a limitation of sampling. The situations in which students are observed are usually opportunistically selected. That is, assessments involve patients who happen to be available at assessment time. Thirdly, it is unclear how much weight should be placed on observations compared to the other ratings during clerkships (orals, written examinations).

Ongoing assessment of practical performance can be improved in a number of ways. Ideally, student observation should be based on a blueprint of problems they must master during clerkships. Standardisation of assessment instruments used to observe students and training of supervising physicians will enhance reliability. Ongoing assessment must be accompanied with a careful report of student performance done by professionals who collaborate with the student. In many cases, this is not the supervising physician, but the residents, nurses or peers. To improve performance these reports must be discussed regularly. Other sources of information are also valuable such as student logbooks, documenting what a student has seen and done on the ward, peer and self-assessments or patient reports. The frequency and timing of the different assessments must be established, as must rules to combine the results of the different assessments. This should lead to an integrated overall comprehensive evaluation system in practice.

The Pyramid of Miller Extended

Miller’s pyramid is useful to describe medical competence in a hierarchical way. It is important not only for medical schools using PBL but also for more traditional schools.

The incremental ambition of PBL is a focus on the responsibilities of the learner for his own learning process and the concept of lifelong learning. It stresses a constructivistic approach to knowledge acquisition, requiring the learner to be actively involved in the process and working in collaboration with others.

To achieve this, an instructional situation should be created in which students are required to collaborate to optimise their learning process such as is done in tutorial groups. Equally important is allowing time to reflect and to discuss student progress. Providing regular and standardised feedback about behavioural performance of students in different educational contexts also fits very well into the learning process.

If responsibility for the learning process is to be taken seriously, students must master the skills needed to optimise their own learning. Important among these are intrapersonal skills like personal management, time management, stress management and the ability to reflect. Interpersonal skills are equally important. Included here are: effectively dealing with others; communication skills; dealing with feedback, leadership skills and sticking to appointments. These intra and interpersonal skills, together described as “metacognitive” skills, are not originally represented in Miller’s pyramid, but can be easily added as another dimension of the same pyramid. It is clear that these skills are important not only for the learning process but also for future performance.

We will not discuss the different possible layers of competence that are created with a new dimension in the pyramid in relation to metacognitive skills. However, it is important that those metacognitive skills be part of the assessment process in PBL. This may be achieved using the same methods described in previous sessions. For example, ongoing observation will be appropriate to assess how students actually perform in different educational contexts.²⁴ Self reports and reflections on learning will have a more central place in the assessment of these skills.

Another example is the use of tutorial groups to assess and give feedback to students. This can be done with the help of a global rating scale that contains criteria referring to individual performance in the group. Such criteria usually relate to individual performance and collaboration within the group. Halfway through the educational activity, students should get feedback on their performance from their tutors and peers using criteria on the rating scales. Thus, students will have an opportunity to alter their behaviour based on the feedback. At the end of the educational activity, this process is repeated and a mark is given for overall performance. Repeated assessments with feedback in many educational situations will provide a clear view of individual weaknesses and strengths. In PBL,

the authentic setting is not a clinical ward, but a specific educational situation. The examiner is not a supervising clinician but a tutor or a peer. The method (ongoing observation) is the same for both with the same strengths and weaknesses (standardisation and sampling).

Another possible way to assess the learning process is portfolio assessment.²⁵ In portfolio assessment, the student regularly provides the teacher information about his competence. The responsibility for gathering the information remains with the student. Teachers determine what aspects of competence must be represented in the portfolio. Several times during the academic year the student and teacher meet and during the meetings, the student reflects on his or her performance. In discussion with the teacher, weak and strong points will be identified and discussed. Student and teacher have to agree upon measures to improve weak points. At the end of the year, the teacher (or committee) can decide to let the student pass or may give a conditional pass. This means that the student has to bring about more evidence before he is promoted to another year. Of course the student can fail but this should not be a surprise when the student and his teacher have met regularly. The active role of the student in this type of assessment fits very well into the framework of PBL.

Conclusions and Suggestions

Research in assessment and the models used in assessment tells us the following:

- No ideal assessment instrument exists for PBL
- All assessment instruments are fraught with problems
- Successful student assessment (evaluation) requires the use of several complementary instruments

Whichever system is used, it is critical to have a match between the system of assessment and the educational goals. In PBL, students learn to apply knowledge and skills in simulated and real life medical contents. Therefore, the system selected must focus on the higher cognitive abilities represented in the upper layers of Miller’s pyramid. Maintaining this focus requires use of appropriate, relevant content in the questions. Since assessment has such a significant impact on student learning, it is useful to develop instruments, which do not encourage last minute student preparation for the test. In our opinion, the progress test described previously achieves this goal. Alternatively, one can harness the positive energy generated by the spectre of assessment and use it to support the educational process. This can be done in tutorial groups by having students write a paper about a learning goal, do a presentation individually or in groups or prepare and role-play a relevant patient problem. In addition, ongoing assessment of clinical ward performance may be useful. Metacognitive skills assessment is important. Active participation in group activities,

collaboration with colleagues and assumption of responsibility are examples of dimensions subsumed in this domain. Ongoing observation of students will allow assessment of these important areas. The currently popular method of portfolio assessment fits very well into the learning structure of PBL. It puts the responsibility for assessment onto the shoulders of the student and encourages reflection on strengths and weaknesses.

REFERENCES

1. Van der Vleuten C P M. The assessment of professional competence: theoretical developments, research and practical implications. *Adv Health Sci Educ* 1996; 1:41-67.
2. Miller G E. The assessment of clinical skills/competence/ performance. *Acad Med* 1990; 65:63-7.
3. Van der Vleuten C P M, Verwijnen G M, Wijnen W H W. Fifteen years of experience with progress testing in a problem-based curriculum. *Med Teacher* 1996; 18:103-10.
4. Rimoldi H J A. The test of diagnostic skills. *J Med Educ* 1961; 36: 73-9.
5. McGuire C H, Babott D. Simulation technique in the measurement of problem solving skills. *J Educ Measurement* 1967; 4:1-10.
6. Hodgkin K, Knox J D E. Problem centered learning: The modified essay question in medical education. Edinburgh: Churchill-Livingstone, 1975.
7. Barrows H S, Tamblyn R M. The portable patient problem pack (P4). A problem-based learning unit. *J Med Educ* 1977; 52:1002-4.
8. Elstein A S, Shulmann L S, Sprafka S A. Medical problem solving: An analysis of clinical reasoning. Cambridge, MA: Harvard University Press, 1978.
9. Newble D I, Hoare J, Baxter A. Patient management problems: Issues of validity. *Med Educ* 1982; 16:137-42.
10. Norcini J J, Swanson D B, Grosso L J, Shea J A, Webster G D. Reliability, validity and efficiency of multiple choice questions and patient management problem item formats in the assessment of physician competence. *Med Educ* 1985; 19:238-47.
11. Bordage G, Page G. An alternative approach to PMPs: The "key features" concept. In: Hart I R, Harden R M, editors. *Further Developments in Assessing Clinical Competence*. Montreal: Heal-Publications, 1987:59-75.
12. De Graaff E, Post G, Drop M. Validation of a new measure of clinical problem solving. *Med Educ* 1987; 21:213-8.
13. Ward W C. A comparison of free-response and multiple-choice forms of verbal aptitude tests. *Appl Psychological Measurement* 1982; 6:1-11.
14. Schuwirth L W T S. An approach to the assessment of medical problem solving: computerized case-based testing. Maastricht: University Press, 1998.
15. Schuwirth L W T S, Blackmore D E, Mom E, Van den Wildenberg F, Stoffers H E J H, Van der Vleuten C P M. How to write short cases for assessing problem-solving skills. *Med Teacher* 1999; 21:144-50.
16. Harden R, Stevenson M, Downie W, Wilson G. Assessment of clinical competence using objective structured examinations. *BMJ* 1975; 1:447-51.
17. Stillman P, Swanson D. Ensuring the clinical competence of medical school graduates through standardized patients. *Arch Intern Med* 1987; 147:1049-52.
18. Newble D, Jeager K. The effect of assessments and examinations on the learning of medical students. *Med Educ* 1983; 17:165-71.
19. Carpenter J L. Cost analysis of objective clinical examinations. *Acad Med* 1995; 70:828-33.
20. Poenaru D, Morales D, Richards A, O'Connor H M. Running an objective structured clinical examination on a shoestring budget. *Am J Surg* 1997; 173:538-41.
21. Vu N V, Barrows H S. Use of standardized patients in clinical assessments: recent developments and measurement findings. *Educ Res* 1995; 23: 23-30.
22. Van Luijk S J, Van der Vleuten C P M. A comparison of checklists and rating scales in performance-based testing. In: Hart I R, Harden R M, Des Marchais J, editors. *Current Developments in Assessing Clinical Competence*. Montreal: Can-Heal, 1992:357-62.
23. Van der Vleuten C P M, Scherpbier A J J A, Dolmans D H J M, Schuwirth L W T, Verwijnen G M, Wolfhagen H A P. Clerkship assessment assessed. *Med Teacher* 2000; 22:592-600.
24. Van Luijk S J, Smeets J G E, Smits J, Wolfhagen I, Perquin M L F. Assessing professional behaviour and the role of academic advice at the Maastricht Medical School. *Med Teacher* 2000; 22:168-72.
25. Snadden D, Thomas M. The use of portfolio learning in medical education. *Med Teacher* 1998; 20:192-9.