Assessment in Problem-Based Learning Medical Schools: A Literature Review

Mathieu R. Nendaz and Ara Tekian
Department of Medical Education
University of Illinois College of Medicine
Chicago, Illinois, USA

Background: *Despite widespread use of problem-based learning (PBL) in medical schools, no review currently exists on its assessment. Given the importance of assessment for any curriculum, a critical review of the literature was conducted to explore whether the assessment methods match the philosophical tenets of PBL.*
Summary: *Articles from MEDLINE and other databases on the assessment of PBL were reviewed. The following areas require special attention by PBL medical schools: enhancement of formal continuous formative assessment; use of the context of a working problem to assess knowledge and problem-solving skills; prevention of negative steering effects by a judicious choice of assessment content, instruments, and timing; and implementation of a longitudinal and centralized student profile.*
Conclusion: *Despite the existence of general practical recommendations on assessment in PBL settings, this review reveals a lack of uniformity and consensus on the practical application of general principles. Topics for future research are highlighted.*

Teaching and Learning in Medicine, *11*(4), 232–243

Problem-based learning (PBL) in small groups is an instructional method designed to respond to a number of concerns about lecture-based and fact and information-based curricula. The latter have been shown to induce passive learning, rote memorization and forgetting, as well as insufficient skills to maintain competency over time in the context of a continually growing knowledge base in medical sciences.[1] PBL is defined as "the learning that results from the process of working toward the understanding or resolution of a problem" (p. 18).[2] According to Barrows, four main objectives must be achieved for a truly problem-based approach: (a) structuring knowledge for better recall and application in clinical contexts, (b) developing an effective clinical reasoning process, (c) developing self-directed learning, and (d) increasing motivation for learning.[3] Although PBL does not necessarily include self-directed learning, Barrows' taxonomy will be used in this article, in which true PBL meets all four criteria.[3]

PBL has been implemented in many medical schools around the world, and reviews have been published on its outcomes in the basic sciences[4–6] and in the clinical setting.[7] Although recommendations and guidelines pertaining to student assessment in PBL have been published,[1,2,8–10] whether and how these principles are applied in practice is difficult to determine in the absence of any comprehensive review on this topic. Given the impact of any assessment on student learning,[11] striving to obtain a match between assessment procedures and the curricular tenets of PBL is essential. This work focuses on the research literature published on assessment in PBL curricula, with the following goals: Report the published assessment experiences of PBL medical schools, analyze how recommendations are fulfilled, and highlight potential areas to be improved or investigated. Unlike a survey, this review does not necessarily totally reflect what every PBL medical school is actually doing now, but rather shows what they have studied and published.

The literature review was based on a computerized search of the following databases: MEDLINE, the Educational Resources Information Centre (ERIC), and Current Contents, from 1966 to 1998, using the key-words "problem-based" or "case-based learning" and "assessment" or "evaluation" in medical education. Research was limited to articles in the English

We thank Professors Richard Foley, PhD, and Reed G. Williams, PhD, for their helpful comments on previous versions of this article.
Correspondence may be sent to Ara Tekian, PhD, MHPE, International Programs, Department of Medical Education (MC 591), University of Illinois College of Medicine, 808 South Wood Street 986 CME, Chicago, IL 60612–7309. E-mail: Ara.Tekian@uic.edu

language and to the field of human medicine. Manual searches of the proceedings of the Association of American Medical Colleges conferences on Research in Medical Education, as well as of the proceedings of the Ottawa Conferences on Medical Education were also conducted. Published books on PBL and the bibliographies of retrieved articles were reviewed to identify additional relevant sources. Journal articles and book chapters were selected if they addressed the issue of the assessment in PBL medical schools. Only medical schools fulfilling Barrows' criteria of the PBL taxonomy[3] were included.

Among 94 citations found, 19 dealt with general aspects of assessment in PBL, 60 addressed assessment issues in the context of PBL applied totally or partially to the curriculum, and 15 were related to PBL applied to a single discipline. Thirty-two medical programs were represented:

1. Eleven in which PBL was the predominant instructional method for all students (43 references).
2. Five in which PBL was the predominant instructional method for a parallel track applied separately from a conventional curriculum (12 references).
3. Three in which PBL was mixed with features of a conventional curriculum (five references).
4. Seven in which problem-based approaches were applied to a specific field in preclinical years (seven references).
5. Six in which problem-based approaches were applied to a specific field in clinical years (eight references).

For each theme pertaining to assessment in PBL, we report the general recommendations from the literature and analyze how these recommendations are applied in practice, based on the published works.

Assessment as a Formative Process

Formative assessment is one of the main features of PBL. It must be a continuous process, integrated with learning, that favors self-assessment[2] and assists students in defining their own learning needs.[12]

Detailed descriptions of formative assessment were poorly represented among the articles. Thirteen programs[13–25] essentially described their summative processes, where periodic assessment may influence each student's progress in various ways. Only two medical schools provided an extensive description of a formal formative assessment. McMaster University provides a program including predominantly formative assessment with no summative or end-of-course examinations.[26] Personal goals are set by students with their tutors and formative assessment permits an oriented remediation that may be proposed by the faculty or identified by the students themselves (self-referral).[26,27]

In 1994, the Newcastle University Medical School[12] introduced a formal formative assessment program that incorporates all the instruments used for later summative purposes. Faculty members are committed to mark students' activities, and this assessment encompasses a written examination on understanding and knowledge of the basic and clinical sciences, group function, clinical reasoning, self-directed learning, and communication and interaction skills.

The low number of reports describing formative assessment in detail does not necessarily mean that medical schools do not provide it, but may reflect the difficulty in making formative assessment a formal process. First, the implementation and maintenance of formal formative assessment demand additional involvement of staff members and may require additional resources.[12] Second, when PBL and traditional tracks are running simultaneously in an institution, there may be internal pressure to demonstrate the effectiveness of PBL through summative tests.[15,28] Third, it is not clear whether student outcomes are better when formal formative assessment is conducted. This issue needs to be explored more extensively.[12]

Content of Assessment and Measurement Instruments

A central role of PBL is the "acquisition and nurturing of problem-solving skills to be used for a wide variety of clinical problems."[9] Attempts have been made to measure these skills independently of knowledge, but there is little to support the hypothesis that problem-solving skills are really independent of the knowledge context in which the problem-solving process takes place.[29] PBL leans more toward acquiring integrated knowledge in the context in which it will be used later.[30,31] Therefore, assessment in PBL must challenge traditional assessment systems.[2,15,32,33] It should focus less on the content, recall of isolated facts, and outcomes and focus more on the processes and performances in context, because the fact that students know something is no assurance that they know when or how to use that information.

The content of tests often represents values reinforced by faculty, and assessment exerts an important influence on the curriculum (steering effect),[11] including problem-based curricula.[34] Assessment may invite students to "study for the test" and reinforce rote memorization if poorly designed.[35] Such attitudes jeopardize the spirit of PBL and self-directed learning. Therefore, assessment must be consistent with how students learn in PBL settings,[9] match the educational objectives of the program, and relate to the objectives set by students. To remain student centered, Barrows[8]

suggested using the learning issues developed by each group during the tutorials to direct the examination content. Merely listing detailed behavioral assessment objectives would be the antithesis of problem-based self-directed learning by making the assessment teacher centered.[9]

The following sections describe various approaches reported by PBL settings to measure knowledge and problem-solving skills, practical skills, and professional attitudes. These instruments are categorized, according to Swanson's classification,[10] as *outcome-oriented* and *process-oriented* instruments.

Knowledge and Problem-Solving Skills

Outcome-Oriented Instruments

Outcome-oriented instruments assess the product of students' performances rather than how results were achieved. As Barrows[2] emphasized, these measures do not refer to the outcomes of patients.

Multiple-choice questions, short-answer questions, and extended-matching items. Multiple-choice questions (MCQs), which are also broadly used in PBL, have been criticized for assessing isolated and independent facts and testing only recall.[15] If the task posed for the examinees is appropriate, however, this method has been shown to reflect not just recall of facts, but also problem-solving skills.[10] This is a controversial issue in the PBL field. Barrows[8] proposed that, if such test formats are used, they should be in context, immediately following work with a patient problem (standardized patient or computer simulation). Guessing and "cueing" are other criticisms typically associated with MCQs. Increasing the number of relevant response options, however, is a straightforward way to address these issues. From a psychometric standpoint, MCQs offer high internal consistency reliability, because this format allows for broad sampling of content domains, and high validity if they are constructed appropriately.[10]

Guessing and "cueing" are less problematic with short-answer questions (SAQs), because no choices are offered to the examinee. Potential problems with this format are their lack of precision regarding what is expected from the students, the increased faculty time and effort for scoring, and lower test reliability due to the fewer number of test items and more subjective scoring.

The extended-matching item format, where the examinee must choose among numerous alternatives, may represent a compromise between MCQs and SAQs. In addition to reducing the risk of guessing and "cueing," this format offers a more objective and automated scoring process.

The Progress Test. To minimize the negative steering effect of assessment on students' learning, the University of Maastricht (formerly University of Limburg) implemented the Progress Test (PT), which is not linked to any particular curricular block.[36] It reflects the end objectives of the curriculum and samples knowledge across all disciplines and content areas in medicine relevant for a medical degree. Questions are at all taxonomic levels, provided that the knowledge addressed is functional for the student on graduation. Each test consists of about 250 true–false items stratified across 15 categories based on the International Classification of Diseases. It is administered four times a year to all students in the medical school. Results are used in formative and summative ways. In examining the validity of the PT with regard to the growth of clinical reasoning skills, Boshuizen and coworkers[37] found a high correlation ($r = .93$) with a test on clinical reasoning. They inferred that the PT was a valuable tool for monitoring students' progress and that true–false questions may assess more than factual knowledge.

The PT was also implemented at McMaster University in 1992,[27] as a 180-item, multiple-choice test administered in 3-hr sittings to all classes in the medical school three times a year. Test–retest reliability over successive administrations was .53 to .64, and predictive validity of the cumulative score was in the order of .60. The overall impact of the test on the students' learning habits was minimal, and 73% of the students mentioned that the test had no effect on tutorial-group function. Results are used formatively, which allows students to identify their needs for remediation and permits identification of students who have severe and persistent problems.

Essay exams. The advantages of essay exams are that they can more easily reflect the problem-solving skills of students and provide a format that encourages students to integrate their knowledge. Drawbacks include the risk of asking questions that are too vague and the low reliability due to the small number of questions and subjective scoring. To limit these problems, a larger number of short essays are better than a few longer questions.[10] The use of essay questions was reported by few PBL medical schools.[16,22,38]

Oral and structured oral examinations. The advantages and disadvantages of oral examinations are similar to essay exams, but additional logistic and scoring problems may be encountered. To improve their objectivity, attempts have been made to standardize the examinations by preparing sets of hypothetical cases and suggesting lines of questions and scoring criteria in advance.[39] The use of oral assessment was commonly reported by PBL medical schools.[24,28,40–48]

Patient-management problems,[49] computer simulations, and paper cases. To test clinical reasoning and problem-solving by case simulation, patient-management problems were used extensively until the late 1980s, when evidence of psychometric flaws led to their discontinuation.[10,39] The field of computer simulation holds greater promise. High-fidelity models of patient care, which present diseases in a dynamic way and include realistic pictures of patient findings, are anticipated.[50–53] Assessment through computer simulations was reported by two surgical programs using PBL in a clinical setting.[54,55]

In an attempt to find an inexpensive, standardized, and easily administered and scored method of assessment, paper cases were developed in the setting of a course entitled "Issues in Contemporary Medicine" given during the 2nd year at the University of Illinois College of Medicine at Rockford. As described by Hand,[56] this method tested patient–physician communication, with emphasis on patient compliance. Questions were related to a clinical situation given in a sequential format to be responded to in an open-ended fashion. To account for multiple correct responses, the answer key contained alternatives requesting the personal judgment of the rater. No psychometric properties were reported. Other formats of patient management assessment reported by PBL programs include direct observation of students,[16,19,54,57–62] log-books recording experiences and skills performed,[16,23] case descriptions,[23,63] and write-ups.[16,57,64]

Modified essay question.[65] To test the clinical reasoning process in a PBL setting, the faculty of medicine, University of Newcastle, adapted and used the modified essay question (MEQ), a serial, structured question examination presenting clinical cases in sequence in a booklet.[65] To simulate the finite temporal sequence of decision making in clinical practice, the MEQ does not permit students to preview items by looking forward in the booklet, or to come back to change decisions on earlier items. The test evaluates knowledge in clinical problem solving and practical skills, such as interpretation of radiographs, electrocardiograms, microscopic sections, or videotaped behavior segments. Answers are given in an open format. The test is administered at the end of each term, and each item is marked "satisfactory" or "unsatisfactory" in accordance with pre-established criteria (mandatory level of competence). Students with questionable results have extensive oral interviews with evaluators. Feletti[65] reported that the MEQ had globally acceptable reliability (Cronbach's coefficient α = .57–.91) and construct validity, and was invaluable for diagnosing students' weaknesses in clinical problem solving as well as for providing a concrete basis for remediation. A comparison between the MEQ, which has a fixed progression though the problem, and a structured oral assessment allowing student-initiated progression did not determine which instrument provided a better measure of clinical problem solving. Each format offered separate evidence of validity and reliability.[40] Eight years after its implementation, an evaluation of the use of the MEQ[66] revealed a decrease of items testing problem solving in comparison with recall or interpretation items, highlighting the need to regularly check the quality of the MEQ item bank. The use of the MEQ has also been reported by other PBL medical schools.[14,24,25,32,46,61,67–70]

Clinical reasoning test modules.[71] Clinical reasoning test modules were developed at Southern Illinois University from the PBL Module[72] to reflect the hypothetico-deductive model of reasoning. Students were asked to compare and contrast a set of hypotheses by eliciting pertinent information from history, physical findings, and ancillary tests. Students received the same problem in the form of multiple workbooks and were asked to carry out an inquiry strategy and decide on initial hypotheses and tests to perform to reach a final hypothesis. This step tested the reasoning process and was corrected through a pre-established score. Self-directed study skills were assessed through a second step where students listed learning issues and resources from the case, corrected their workbooks after self-study, and finally summarized what they had learned and evaluated their own performance.[73] A study by Williams and colleagues[71] reported that the test successfully approximated physician ratings made by direct observation (r = .66–93), could possibly predict clerkship performance (r = .27), and provided an efficient way for assessing clinical reasoning without requiring direct observation by physicians.

Clinical reasoning exercise.[74] The clinical reasoning exercise was developed at McMaster to address the need for testing students on 10–20 cases to ensure reliability. It consists of an oral or written examination containing multiple questions (cases) administered within a 30–60-min period, with only a brief assessment for each case. Neville and coworkers[74] reported acceptable reliability (alpha coefficients ranging from .53 to .64) and a good correlation with objective tests (r = .45–.61).

Problem-analysis questions.[13] The problem-analysis questions, developed and used at the University of Sherbrooke, consist of context-dependent short-answer questions. They were designed to test the students' abilities to "analyze information of a short vignette, generate and evaluate hypotheses, and propose explanations to problems." The internal consistency adjusted for test length was low to moderate (r = .38–.78), probably, in part, because of the variation in scoring these questions. Construct validity was as-

sessed by having the students rate the taxonomic level of questions. Only one third of the questions were judged to be at the analysis level, but more than two thirds were at or beyond the comprehension level.

Standardized patient-based tests. Although standardized patient (SP)-based tests are used most often to measure history taking, physical examination, and interpersonal and communication skills, they also can provide the basis for assessing knowledge and problem solving related to the case through postencounter oral, written, or computerized tests exams.[1,8,75] In this case, the psychometric properties of the test will essentially depend on the postencounter method used.

Although the majority of medical schools reported the use of SPs in their assessment programs, few described encounters with SPs as a basis for assessing applied knowledge and clinical reasoning in the context of a problem. Exceptions are Southern Illinois University[18] and the medical schools using the Individual Process Assessment (IPA)[19,24,76] (cf. Learning Exercises section, which appears later).

In summary, although traditional ways to measure knowledge and problem-solving skills, such as MCQs, SAQs, and oral examinations, are broadly used in PBL medical schools, many attempts have been made to develop tests reflecting the reasoning process of physicians. Although no consensus seems to exist in the reviewed literature on the best method to use, many approaches have their own merit. For example, the PT is helpful to prevent negative steering effects on learning; the MEQ, with carefully constructed items, or the clinical reasoning exercise offer valuable approaches embedding problem solving and content knowledge assessment. Although PBL medical schools commonly report the use of SP-based tests, they rarely mention conducting postencounter examinations. However, use of SP-based tests seems an attractive approach to broadly cover dimensions of competence while providing sound psychometric properties if well-constructed MCQs or extended-matching items are used. The next section reviews the instruments used to assess the processes of PBL.

Process-Oriented Instruments

Process-oriented instruments were developed to reflect how well the educational objectives or criteria advocated by PBL are achieved.[2] These instruments essentially assess the development of problem-solving skills, self-directed learning and self-assessment skills, communication skills, and critical thinking. They include learning exercises and tutor, peer, and self-assessment.

Learning Exercises: The triple jump-based exercises, the individual process assessment, and the four step assessment test (4SAT). In the triple jump (TJ) exercise,[77] students discuss a written clinical scenario and identify the related learning goals, review the learning materials individually, and come back to present their conclusions and judge their own performances. This method aims primarily at assessing problem-solving and self-directed learning skills. Nevertheless, psychometric flaws related to the low number of cases examined limit their use for summative purposes,[10,32] and variations of this test have been developed.

At the University of Hawaii,[25] an attempt was made to increase the objectivity of this test. TJ subjective scores were generated by the faculty members and were derived from responses to pre-established statements about students' performances. TJ objective scores were based on the comparison of pre-established criteria with students' written answers to the first step of the process—the hypothesis-generation step. There was a moderate correlation between TJ subjective and objective scores (r = .53) and a high internal consistency for the TJ as a whole (Cronbach's coefficient α = .89). There was no correlation between TJ scores and other objective tests such as the MEQ or the laboratory practicum. Reliability and validity remained a problem, because students worked on a single problem, and the different problems used were unequal in difficulty. However, as the TJ is graded on a pass–fail format, the authors concluded that a subjective assessment was adept at discerning marginal-to-unsatisfactory performances, and that the objective scores were helpful in corroborating subjective impressions for students who received low subjective scores.

At the University of Linköping,[14,78] the exercise is based on a videotaped interview with a real patient, which is evaluated by a general practitioner and a basic scientist. After an oral assessment of knowledge and understanding, the student formulates a learning task and is given 4 hours to research the information before presenting the newly acquired knowledge to the same examiners in addition to a teacher from the medical library. No psychometric properties were reported for this test.

To test students in their PBL course entitled "Medical Problems," Friedman and colleagues[38] described another variation of the TJ. Cases are discussed in groups with tutors, and students have 1 week to develop individual written reports addressing the three most important learning issues developed by their group and two additional questions given to the entire class. Tutors from different groups grade these five written short-essay responses. Psychometric analyses yielded generalizability coefficients ranging from .52 to .71 and low correlations with tutor ratings and scores in other courses (r = .15–.32).

The IPA,[15,76,79] developed at the University of New Mexico, is designed to assess the full range of activities

expected from a physician. Each student is videotaped during an encounter with an SP and is judged on interpersonal and clinical skills. The students submit write-ups about the pertinent findings, hypotheses, and learning needs, and have 2 days to research the identified learning issues. They then return to discuss learning and scientific issues with two faculty evaluators. The process requires over 50 hr per student and demands extended involvement of faculty members. Interrater reliability was reported as "problematic," but no reliability or validity studies were reported. Other schools reported the use of the IPA in their PBL programs.[19,24]

The 4SAT is a test recently reported by the University of Queensland, Australia,[80] that aims to assess individual knowledge and clinical reasoning, as well as group process. Four steps characterize this process. First, students must solve a case scenario individually and in writing by identifying key features, developing hypotheses, proposing basic mechanisms to explain symptoms, requesting additional clinical data to refine their hypotheses, and identifying learning issues. In step two, the same process is repeated at the group level and new information about the case is presented. Observers score the tutorial process. After a period of self-directed learning (step three), students take a written content exam (step four) that assesses the "top 10" learning issues derived from all groups' answers that were previously posted on an electronic bulletin board. Interrater agreement was greater than 80%, and the 4SAT scores correlated well with scores derived from other objective instruments ($r = .49$).

The Universities of Newcastle[59] and Linköping[46] reported original assessment activities at the end of their curricula. Components of these examinations include self-assessment of one's own learning (Medical Independent Learning Exercise), review of scientific papers, and presentation of a personal scientific work that is reviewed by student peers. This new examination was perceived as a positive experience because it was not merely the repetition of the types of examination used at the end of each semester and it tested other qualities. Furthermore, including mandatory scientific works and readings was thought to allow for the development of a science-based professional attitude.[69] No study of reliability and validity was reported.

Tutor, peer, and self-assessment. Assessment of processes and attitudes during tutorial sessions is thought to embody PBL principles, and is the central focus of student assessment.[2,8,32] Through regular contacts among students and tutors, a climate of mutual knowledge allows the tutor to effectively assess the development of clinical reasoning, self-directed learning, communication, internal motivation, critical thinking, and the ability to work effectively in a team. Although these methods address the major principles of PBL, they possess psychometric shortcomings[81–83] that limit their use in high-stake decision making.

The majority of PBL schools report assessment during tutorials, but its purpose (summative or formative) is usually not obvious. When stated, the use of assessment during tutorials is variable. For example, the University of Maastricht does not use tutorial ratings in summative assessment because the roles of tutor–rater and tutor–teacher are considered to be incompatible.[84] On the other hand, the University of Sherbrooke uses tutorial summative assessment systematically, and developed a comprehensive assessment instrument—the Tutotest. Here, items were generated through the critical incident technique. Tutors rated each student on a 10-point scale for 44 items representing four factors: effectiveness in the group, communication and leadership skills, scientific curiosity, and respect for colleagues. Reliability was high (Cronbach's coefficient $\alpha = .98$), the correlation coefficient with tutor global assessments was .64, and the correlation with students' written examination was .39.[85]

Another method to evaluate students in their group is the Group Task exam developed at Newcastle University.[86] In the first step, two tutors observe a group of students during their first tutorial, where they are given a problem to solve. Several dimensions (dynamic, organization, reasoning, and critical thinking) of the group process are assessed. Judgments are reported as satisfactory or unsatisfactory. Groups considered to be unsatisfactory are required to be re-assessed after a period of remediation. During a second step, after a 24-hour period of self-directed learning, each student is evaluated by an oral examination of the learning goals, resources used, and understanding of the problem. This assessment was used summatively in this institution. No study of reliability and validity was reported.

In summary, although psychometric properties of process-oriented instruments have still to be improved for high-stake decisions, these methods remain an essential component of a PBL assessment program taken as a whole. They are consistent with PBL philosophy, emphasize its educational objectives, and provide a positive steering effect on learning and useful education. Tutor, peer, and self-assessment develop the ability to give and receive feedback and to appraise one's own needs, which are required in the daily activities of a physician. They also allow detection of potential interpersonal problems that would have remained unnoticed otherwise. Diverse variations of the TJ have been described and all have the merit to assess the different steps of the learning process in a PBL environment. The choice of a method for a specific institution may depend on the number and the availability of faculty members, as well as on the internal organization of the curriculum. The IPA may be suitable if resources are sufficient, whereas the 4SAT offers an interesting alternative allowing for its application to large groups.

Practical Skills

The assessment of clinical skills in PBL settings (history taking, physical examination, management, and communication skills) generally relies on the Objective Structured Clinical Examination (OSCE), an approach to test multiple components in one setting in which a variety of methods can be embedded. Typically, practical skills are assessed using SPs. The OSCE is organized in stations, each usually containing one type of tested activity (such as a focused interview or physical examination), a specific procedure, or a short assessment of diagnostic skills and patient management following an SP encounter.[75] One rater per station may suffice, and this individual does not necessarily have to be a physician. Testing time should be at least 3 to 4 hr, and the length of one station may vary from 5 to 20 min. The use of more, shorter stations (typically 20 stations) is better than a few very long stations. Psychometric properties are acceptable if care is taken on test construction, methods to convert examinee behavior into appropriate scores, and test length.[75,87]

An example of application and variation of skill assessment was reported by the University of Maastricht. At the end of each academic year, all students (including the 5th- and 6th-year students in clerkships) are required to take the Skills Test,[58] similar to the OSCE. It consists of direct observation by faculty members of students' technical or clinical skills through 12 stations of 10 to 30 min (a 2-hr exam). Observed skills are divided into four areas: social (interview), physical examination, laboratory, therapeutic. Observers (and sometimes co-observers) make their judgments with the help of a list of criteria (correct/incorrect/incomplete). Reliability and validity were considered satisfactory in Bouhuijs and colleagues' study,[58] given that the test relies completely on observation. The interrater reliability was .85, overall Cronbach's coefficient alpha was greater than .90, and validity studies yielded "encouraging" findings. The Skills Test is, however, costly and requires long tests for psychometric reasons. Moreover, at this university, students' skills are evaluated only once a year for 2 hr, which is minimal. For these reasons, a written test, the Knowledge Test of Skills, was developed and evaluated.[88,89] This examination consists of 238 true–false items and was constructed as a PT. The questions may focus on the cognitive part of skills, such as the interpretation of a clinical finding, or on the procedure itself. A study by Van der Vleuten and others[88] found reliability (generalizability coefficients .89–.95) and convergent validity (true correlation reaching .89) high enough to propose this test as a supplement to performance tests, with the benefits of helping to overcome test length and costs while providing more reliable composite information.

In summary, for practical skill assessment, the experience acquired with the OSCE makes it a valuable tool if care is taken in constructing tasks that reflect clinical practice. If test length and costs are difficult to overcome, the association of performance tests with a written assessment of knowledge of skills may represent an interesting alternative.

To have a sense of the global value of each reported instrument, we attributed rough ratings to six characteristics related to assessment methods used in PBL, which are summarized in Table 1. Because there are no data permitting the attribution of rigorous and comparable scores to each of these attributes, these individual ratings were given through partly subjective judgments based on the data from the literature. These ratings must be considered critically, because the quality of an individual test also depends on how it is constructed and on the general context of the assessment program. For psychometric descriptions, a "low" grade was attributed to coefficients of less than .25, "moderate" to coefficients ranging from .26 to .65, and "high" to coefficients of more than .65. As may be inferred from the wide range of instruments or set of instruments used by PBL medical schools and by the multiple attempts to develop methods measuring processes rather than outcomes, no single choice emerges as most dominant, and the triangulation of diverse methods is necessary to provide acceptable accuracy to high stake decisions. The use of SP-based tests with postencounter use of well-constructed MCQs or SAQs to test applied knowledge and problem solving is an attractive alternative that respects the principles of PBL and also provides sound psychometric properties. A structured approach of assessment during tutorials, such as the Tutotest, may also be of interest, but this remains to be confirmed.

The balance between process-oriented and outcome-oriented tools must be carefully evaluated. Relying mostly on process-oriented methods may respect PBL principles and have a positive influence on students' learning, but their psychometric properties might limit their application in promotion decisions. On the other hand, the extensive use of outcome-oriented instruments, though psychometrically more sound, may jeopardize the spirit of PBL through a negative steering effect if they are poorly designed. A judicious use of various methods seems, therefore, necessary to take into account these different aspects and meet the objectives of every assessment. Moreover, even though the assessment methods have been evaluated individually, as discussed earlier, the value of each of them must be considered not only as an individual tool, but in the context of the integral educational strategy an assessment program offers. For example, within an assessment program, an unreliable assessment method may be accepted because of other important characteristics, such as positive steering effect on learning, and its weaker psychometric properties may be compensated by other instruments used in the program.

The way PBL is implemented into a curriculum may also influence the choice of assessment meth-

Table 1. *Characteristics of Some Instruments Used in PBL Settings*

	Reflects PBL Process	Reflects Medical Practice	Psychometric Properties	Costs	Faculty Time Needed	Ease to Score
SP-Based Tests With Post-Encounter Follow-Up[a]	++	+++	+++[b]	- - -	- -	++
OSCE[a]	+	+++	+++	- - -	- - -	++
MCQs, SAQs, EMIs[b]	+	+	+++	-	-	+++
PT[c]	+	+	+++	-	-	+++
Tutotest[d]	+++	++	++	- -	- - -	++
MEQ	+	++(+)	++(+)	-	- -	+
Computer Simulation[d]	+	++(+)	+ (?)	- -	-	+++
CRTM[e]	+	++(+)	+	- -	-	++
4SAT[d,f]	+++	++	++	- - -	- - -	++
PAQs	+	++	++	- -	- -	++
IPA	+++	++(+)	+	- -	- - -	+
CRE	+	++	++	- -	- -	+
TJ	+++	++(+)	+	- - -	- - -	+
Tutor-, Self-, Peer Assessment	+++	++	+	-	- - -	+
Essays	++	+(+)	+	- -	- - -	+
PMPs[g]	+	++(+)	+	-	- - -	+

Note: PBL = problem-based learning; SP = standardized patient; OSCE = objective structured clinical examination; MCQ = multiple-choice question; SAQ = short answer question; EMI = extended matched items; PT = Progress Test; MEQ = modified essay question; CRTM = Clinical Reasoning Test module; 4SAT = four step assessment test; PAQ = problem-analysis question; IPA = individual process assessment; CRE = clinical reasoning exercise; TJ = triple jump; PMP = patient management problem. Plus signs mean advantages; minus signs mean disadvantages. + (-) = low; ++ (- -) = moderate; +++ (- - -) = high.

[a]Combination involving more than one method. [b]If well-constructed MCQs, SAQs, or EMIs used. [c]Takes steering effect into consideration. [d]Value to be confirmed by further research. [e]Does not require direct observation by physicians. [f]May be applied to large groups. [g]Discontinued.

ods.[15] For example, the potential need to compare a PBL track with a parallel conventional curriculum may lead to the use of more traditional methods.[28] The type of problem-based curriculum also influences the choice of assessment instruments.[10] When students have full responsibility for determining their own learning objectives ("open discovery"), the use of outcome-oriented instruments may be problematic, because each student will pursue a slightly different curriculum and have different learning objectives, as opposed to programs where the objectives are defined by the faculty ("guided discovery").

Elaborating an assessment plan that respects PBL principles, is reliable and valid, and has no negative steering effect remains a challenging task. Much has been done in developing instruments aimed at reflecting the problem-solving process, but more work will be required before reaching a consensus regarding the best instruments to use. Such research would require going beyond the mere concern of measuring whether learning has occurred during the curriculum and addressing prediction of future performance after graduation.

Criteria Setting, Scoring, Grading, and Reporting Results

Few medical schools reported in detail the way criteria are set, scoring is performed, grades are assigned, and results are reported. In PBL, knowing who has to set the criteria for each specific problem may be difficult, because there is rarely only one correct way to reason, and the learning objectives may vary among groups of students. This is especially true for process-oriented instruments. The documentation of "reference performances" has been proposed to help in this respect.[2] The set of criteria for student performance may, for example, be drawn up from the performances of health professionals (physician, nurse, etc.) encountering the cases prepared for student assessment.

Although the use of Nedelsky and Ebel's methods were respectively reported for MCQs and PAQs in a PBL setting,[13] the way in which passing standards are set in the practice of medical PBL schools has still not been reported extensively. The description of score processes—criterion-referenced or norm-referenced—is also rarely reported, with few exceptions.[15,27]

Grading is usually based on a combination of information provided by tutors and tests. However, the role played by the tutor in making the judgment may vary. In some institutions, this role is central,[32,76] whereas in other institutions, tutors are not involved in the summative process.[84] Final results are reported in either a pass–fail[90] or satisfactory–unsatisfactory format,[22,59,91,92] and more rarely as a grade (letter or number).[13,59]

For formative purposes, reports to students often include detailed individual scores and subscores by discipline, with a comparison of the scores of the whole class.[27,36] To propose sounder educational recommendations, the University of Sherbrooke developed Student Longitudinal Performance Profiles, consisting of

data matrices of progress on a yearly basis.[13] These take into account the evolution of performances in time, as assessed through a variety of instruments.

A satisfactory description of how PBL medical schools set criteria and standards and, score, grade, and report the results was clearly lacking in the reviewed papers. The potential heterogeneity of students' learning objectives and the flexible nature of problem-solving processes make setting standards difficult, and more insight into this issue would be helpful. For grading purposes, methods of standardizing determination of a student's profile for promotion decisions should be better developed and described. Following the progress of each student longitudinally would undoubtedly help with the interpretation of individual test results, where the score of each test would not merely be a "stand-alone" value, but one of the many pieces of the students' overall academic profile. Although this centralization of students' performances is theoretically easier in a PBL setting, where multidisciplinarity is emphasized, there is little evidence in the published literature that such follow up commonly occurs. Centralized databases are expected to help in drawing each student's longitudinal characteristics, to make promotion decisions as fair as possible, and to allow appropriate remediation for small but repeated deficiencies, which might have been ignored if only considered individually.

Frequency of Assessment

In essence, formative assessment must be continuous and include self-assessment as an integral part of the learning process.[2] Summative assessment timing, however, should be carefully planned, because too many tests may induce negative steering effects and lead students to finally abandon their own learning goals to study the goals they expect to appear in the test.[10] To test integration and application of knowledge, summative assessment should occur at intervals of some months and cover a set of different topics.

In the majority of PBL medical schools, tests occur at several stages, at the end of units or blocks, at the end of terms or semesters, and at the end of years or of phases. Many patterns and combinations can be found, however, depending on whether the tests are formative or summative and on what is tested (knowledge, skills, or attitudes). Some steps may be formative, whereas other are summative. For example, end-of-block tests at the University of Maastricht are formative, whereas quarterly PTs are summative. Some steps may be optionally formative or summative, as occurs at Northwest Center for Medical Education, Indiana, where a midterm MCQ test is optional, and its results can be kept or dropped by the student for the final grade.[20] Occurrence of tests may depend on what they measure: every 4[13,79] to every 10[19, 24] weeks for summative tutorial assessments, two or three times a year for knowledge or skills,[13,60,93–95] or even once a year for skills.[22,58,88,96]

No uniformity in test timing arose from the reviewed papers, in part because of the diversity in curriculum organization and assessment instruments used. The lack of description of formative assessment prevents evaluation of its continuity. Additional research is needed to determine the best pace of summative assessment that would prevent the negative steering effect and preserve fair and thorough reflection of students' abilities. In this respect, at least for the assessment of cognitive aspects, the PT may represent a useful alternative.

Conclusion

One must first acknowledge the huge effort that is continually being made by many PBL medical schools to describe and publish their overall assessment plan and evaluate its accuracy. This review revealed essentially the absence of uniformity and consensus on many aspects of assessment, despite the existence of general practical recommendations. Because this work was not based on a survey of the assessment methods used at PBL medical schools but on a review of the published literature, it does not necessarily reflect what medical schools are doing currently. Instead, it reflects what the schools have studied and reported. This could account partially for the lack of uniformity perceived. Additionally, such uniformity may not be expected in the practical application of PBL principles, due to different curricular organization and diverse external pressures. Nevertheless, the following fields appear to deserve special attention: enhancement of formal continuous formative evaluation; use of the context of a working problem to assess knowledge and problem-solving skills; prevention of negative steering effects by the judicious choice of assessment content, instruments, and timing; and the implementation of a longitudinal and centralized student profile. Despite the large range of assessment methodologies used in PBL settings, no single choice emerges, and the triangulation of diverse instruments is required to obtain a fair judgment about students. Moreover, the choice of an instrument should depend not only on its individual properties, but also on the characteristics it may bring to the global instructional value of an assessment plan.

The following research topics surfaced as a result of this review. First, because this review does not necessarily represent the current practice in PBL medical schools (because it is based on published articles within a large space of time), a survey of PBL medical schools would be useful to obtain a cross-sectional picture of the present assessment characteristics. This would bring insight into the lack of uniformity and of

consensus across schools that was perceived in this work. Second, although formative assessment is one of the main principles of PBL, the actual influence of formal formative assessment on student outcomes is unclear, and its relative impact when associated with summative assessment still needs to be defined. Because many PBL medical schools seem to lack in formal formative assessment, as defined by Rolfe,[12] a better determination of its importance is crucial to give this demanding process a chance to be implemented more extensively. Third, it appears that the assessment of problem-solving skills must occur in the context of a clinical problem, yet no unanimity exists on the methods to use. Many attempts have already been made to develop instruments reflecting the reasoning process, and future efforts should probably address the value of different combinations of existing instruments rather than the development of new methods. Fourth, because the heterogeneity of students' learning objectives and the flexible nature of problem-solving processes make setting standards difficult, methods to standardize the determination of students' profiles for promotion decisions should be better developed and described, and the value of centralization of students' performances assessed. Finally, outcomes of PBL students are frequently measured and compared to "traditional" students. Yet, given the importance of the steering effect of assessment on students' learning, comparing diverse instructional methods without taking into account the assessment plans used may lead to actually comparing the impact of assessment methods rather than instructional methods. Any future study comparing instructional methods should, therefore, also consider the facets of assessment, the influence of which extends to every aspect of the curriculum.

References

1. Barrows HS. Problem-based, self-directed learning. *Journal of American Medical Association* 1983;250:3077–80.
2. Barrows HS, Tamblyn RM. *Problem-based learning: An approach to medical education.* New York: Springer, 1980.
3. Barrows HS. A taxonomy of problem-based learning methods. *Medical Education* 1986;20:481–6.
4. Albanese MA, Mitchell S. Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine* 1993;68:52–81.
5. Berkson L. Problem-based learning: Have the expectations been met? *Academic Medicine* 1993;68:S79–88.
6. Vernon DT, Blake RL. Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine* 1993;68:550–63.
7. Foley RP, Polson AL, Vance JM. Review of the literature on PBL in the clinical setting. *Teaching and Learning in Medicine* 1997;9:4–9.
8. Barrows HS. *Practice-based learning: Problem-Based Learning Applied to Medical Education.* Springfield, Illinois: Southern Illinois University School of Medicine, 1994.
9. Norman GR. What should be assessed? In D Boud, G Feletti (Eds.), *The challenge of problem-based learning* (pp. 254–9). New York: St Martin's Press, 1991.
10. Swanson DB, Case SM, van der Vleuten CPM. Strategies for student assessment. In DBG Feletti (Ed.), *The challenge of problem-based learning* (pp. 260–73). New York: St Martin's Press, 1991.
11. Newble D, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Medical Education* 1983;17:165–71.
12. Rolfe I, McPherson J. Formative assessment: How am I doing? *The Lancet* 1995;345:837–9.
13. Des Marchais JE, Vu NV. Developing and evaluating the student assessment system in the preclinical problem-based curriculum at Sherbrooke. *Academic Medicine* 1996;71:274–83.
14. Foldevi M, Svedin CG. The Linköping curriculum: The phase examination in general practice. *Medical Education* 1996;30:326–32.
15. West DA, Umland BE, Lucero SM. Evaluating student performance. In A Kaufman (Ed.), *Implementing problem-based medical education* (pp. 144–63). New York: Springer, 1985.
16. Hamad B. Problem-based education in Gezira, Sudan. *Medical Education* 1985;19:357–63.
17. Vu NV, Bader CR, Vassalli JD. The redesigned undergraduate medical curriculum at the University of Geneva. In Scherpbier A, van der Vleuten C, Tethans J (Eds.), *Advances in medical education* (pp. 532–5). Dordrecht, The Netherlands: Kluwer, 1997.
18. Vu NV, Barrows HS, Marcy ML, Verhulst SJ, Colliver JA, Travis T. Six years of comprehensive, clinical, performance-based assessment using standardized patients at the Southern Illinois University School of Medicine. *Academic Medicine* 1992;67:42–50.
19. Philp JR, Camp MG. The problem-based curriculum at Bowman Gray School of Medicine. *Academic Medicine* 1990;65:363–4.
20. Sivam SP, Iatridis PG, Vaughn S. Integration of pharmacology into a problem-based learning curriculum for medical students. *Medical Education* 1995;29:289–96.
21. Office of Educational Development Harvard Medical School. The New Pathway to general medical education at Harvard University. *Teaching and Learning in Medicine* 1989;1:42–6.
22. Mann KV, Kaufman DM. A response to the ACME-TRI report: The Dalhousie problem-based learning curriculum. *Medical Education* 1995;29:13–21.
23. *ABCD.* Maastricht, The Netherlands: Office for International Relations, Faculty of Medicine, University of Maastricht, 1996.
24. Bradley E, Smith I, Wise C. The problem-based curriculum of the College of Medicine: Medical University of South Carolina. *Twentieth Annual Session: Innovations in Medical Education Exhibits* (p. 143). Washington, DC: AAMC, 1995.
25. Smith RM. The triple-jump examination as an assessment tool in the problem-based medical curriculum at the University of Hawaii. *Academic Medicine* 1993;68:366–72.
26. Neufeld VR, Woodward CA, MacLeod SM. The McMaster MD program: A case study of renewal in medical education. *Academic Medicine* 1989;64:423–32.
27. Blake JM, Norman GR, Keane DR, Mueller CB, Cunnington J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: Psychometric properties and effect on learning. *Academic Medicine* 1996;71:1002–7.
28. Goodman LJ, Brueschke EE, Bone RC, Rose WH, Williams EJ, Paul HA. An experiment in medical education: A critical analysis using traditional criteria. *Journal of American Medical Association* 1991;265:2373–6.
29. Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: An Analysis of Clinical Reasoning.* Cambridge, MA: Harvard University Press, 1978.

30. Norman GR, Smith EKM, Powles ACP, Rooney PJ, Henry NL, Dodd PE. Factors underlying performance on written tests of knowledge. *Medical Education* 1987;21:297–304.
31. Norman G. Reliability and construct validity of some cognitive measures of clinical reasoning. *Teaching and Learning in Medicine* 1989;1:194–9.
32. Neville A. Student evaluation in problem-based learning. *Pedagogue* 1995;5:2–7.
33. Abu-Zidan FM. The international conference on problem-based learning in higher education, September 24–27, 1995, Linköping, Sweden. *Medical Education* 1997;31:390–3.
34. Blumberg P, Daugherty S. *Good student or good physician: What are we encouraging?* Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA, 1989.
35. Frederiksen N. The real test bias: Influences of testing on teaching and learning. *American Psychologist* 1984;39:193–202.
36. van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with Progress Testing in a problem-based learning curriculum. *Medical Teacher* 1996;18:103–9.
37. Boshuizen HP, van der Vleuten CP, Schmidt HG, Machiels-Bongaerts M. Measuring knowledge and clinical reasoning skills in a problem-based curriculum. *Medical Education* 1997;31:115–21.
38. Friedman CP, Murphy GC, Smith AC, Mattern WD. Exploratory study of an examination format for problem-based learning. *Teaching and Learning in Medicine* 1994;6:194–8.
39. Swanson DB, Norman GR, Linn R. Performance-based assessment: Lessons from the health professions. *Educational Researcher* 1995;24:5–11,35.
40. Feletti GI, Gillies AH. Developing oral and written formats for evaluating clinical problems-solving by medical undergraduates. *Journal of Medical Education* 1982;57:874–6.
41. Chamberland M, Des Marchais JE, Charlin B. Carrying PBL into the clerkship: A second reform in the Sherbrooke curriculum. *Annals of Community-Oriented Education* 1992;5:235–47.
42. Engel CE, Clarke RM. Professional education for capability and change. *Higher Education Review* 1986;18:27–35.
43. Palmer JW, Foley RP, Tissot RG, Cerchio GM. Developing and implementing a "basic science clerkship" for first-year students. *Academic Medicine* 1992;67:477–9.
44. Romzick TM, Smith RW. Applying problem-based learning theory to the clinical clerkship. *Academic Medicine* 1990;65:662.
45. Colby KK, Almy TP, Zubkoff M. Problem-based learning of social sciences and humanities by fourth-year medical students. *Journal of Medical Education* 1986;61:413–5.
46. Hammar ML, Forsberg PM, Loftås PI. An innovative examination ending the medical curriculum. *Medical Education* 1995;29:452–7.
47. Rendas AB, Pinto PR, Gamboa T. Problem-based learning in pathophysiology: Report of a project and its outcome. *Teaching and Learning in Medicine* 1998;10:34–9.
48. Bergdahl B, Ludvigsson J, Koch M, Wessman J. Undergraduate medical education in Sweden: A case study of the Faculty of Health Sciences at Linköping University. *Teaching and Learning in Medicine* 1991;3:203–9.
49. McGuire C, Solomon C. *Construction and use of written simulations.* Chicago: The Psychological Corporation, 1976.
50. Friedman CP, France CL, Drossman DD. A randomized comparison of alternative formats for clinical simulations. *Medical Decision Making* 1991;11:265–72.
51. Friedman CP. Anatomy of the clinical simulation. *Academic Medicine* 1995;70:205–9.
52. Woolridge N. The C-ASE Project: Computer-assisted simulated examination. *Journal of Audiovisual Media in Medicine* 1995;18:149–55.
53. Hoffman H, Vu D. Virtual reality: Teaching tool of the twenty-first century? *Academic Medicine* 1997;72:1076–81.
54. Schwartz RW, Burgett JE, Blue AV, Donnelly MB, Sloan DA. Problem-based learning and performance-based testing: Effective alternatives for undergraduate surgical education and assessment of student performance. *Medical Teacher* 1997;19:19–23.
55. Schuwirth L, van der Vleuten C, van den Wildenberg F. *A computerized surgery examination using short cases.* Paper presented at the 8th Ottawa International Conference on Medical Education. Philadelphia, PA: National Board of Medical Examiners, 1998.
56. Hand JD. Problem-based paper cases for evaluating students in "issues in contemporary medicine." In PAJ Bouhuijs, HG Schmidt, HJM Van Berkel (Eds.), *Problem-based learning as an educational strategy* (pp. 229–38). Maastricht, The Netherlands: Network Publications, 1993.
57. Neufeld VR, Barrows HS. The "McMaster philosophy:" An approach to medical education. *Journal of Medical Education* 1974;49:1040–50.
58. Bouhuijs PA, van der Vleuten CP, van Luyk SJ. The OSCE as a part of a systematic skills training approach. *Medical Teacher* 1987;9:183–91.
59. Feletti GI, Saunders NA, Smith AJ, Members of the Assessment and Phase V subcommittees of the Undergraduate Education Committee. Comprehensive assessment of final-year medical student performance based on undergraduate programme objectives. *The Lancet* 1983;2:34–7.
60. Huber P, Perrier A. A preclinical practice skills program. *Academic Medicine* 1997;72:432–3.
61. Hassan F, Ezzat E, Faris R, Fam R. The development of a valid student assessment system in community-based medical schools. In PAJ Bouhuijs, HG Schmidt, HJM Van Berkel (Eds.), *Problem-based learning as an educational strategy* (pp. 249–58). Maastricht, The Netherlands: Network Publications, 1993.
62. Goldstein DA, Hoffman KI, Bethune J. The role of the student ward in the medical clerkships. *Journal of Medical Education* 1985;60:524–9.
63. McLeod PJ, Whittemore NB. A problem-based clinical course in general internal medicine. *Medical Teacher* 1989;11:169–75.
64. Echt R, Chan S-W. A new problem-oriented and student-centered curriculum at Michigan State University. *Journal of Medical Education* 1977;52:681–3.
65. Feletti GI. Reliability and validity studies on modified essay questions. *Journal of Medical Education* 1980;55:933–41.
66. Feletti GI, Smith EK. Modified essay questions: Are they worth the effort? *Medical Education* 1986;20:126–32.
67. Pallie W, Carr DH. The McMaster medical education philosophy in theory, practice and historical perspective. *Medical Teacher* 1987;9:59–71.
68. Foldevi M, Trell E. Learning the basics of medicine in general practice in the Faculty of Health Sciences, Linköping, Sweden. *Annals of Community-Oriented Education* 1993;6:97–113.
69. Dahle LO, Forsberg P, Svanberg-Hard H, Wyon Y, Hammar M. Problem-based medical education: Development of a theoretical foundation and a science-based professional attitude. *Medical Education* 1997;31:416–24.
70. Schwartz RW, Donnelly MB, Nash PP, Young B. Developing students' cognitive skills in a problem-based surgery clerkship. *Academic Medicine* 1992;67:694–6.
71. Williams RG, Vu NV, Barrows HG, Verhulst S. Profile of the clinical reasoning test (CRT): An objective measure of problem solving skills and proficiency in using medical knowledge. In HG Schmidt, MV de Valdes (Eds.), *Tutorials on problem based learning* (pp. 81–90). Assen, The Netherlands: Van Gorcum, 1983.
72. Distlehorst LH, Barrows HS. A new tool for problem-based, self-directed learning. *Journal of Medical Education* 1982;57:486–8.
73. Barrows HS. *How to design a problem-based curriculum for the preclinical years.* New York: Springer, 1985.

74. Neville AJ, Cunnington J, Norman GR. Development of clinical reasoning exercises in a problem-based curriculum. *Academic Medicine* 1996;71:S105–7.
75. van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: State of the art. *Teaching and Learning in Medicine* 1990;2:58–76.
76. Kaufman A, Mennin S, Waterman R, et al. The New Mexico experiment: Educational innovation and institutional change. *Academic Medicine* 1989;64:285–94.
77. Painvin C, Neufeld V, Norman G, Walker I, Whelan G. The "triple jump" exercise: A structured measure of problem-solving and self-directed learning. In *Proceedings of the 18th Annual Conference on Research in Medical Education* (pp. 73–7). Washington DC: AAMC, 1979.
78. Foldevi M, Sommansson G, Trell E. Problem-based medical education in general practice: experience from Linköping, Sweden. *British Journal of General Practice* 1994;44:473–6.
79. Kaufman A, Klepper D, Obenshain SS, et al. Undergraduate medical education for primary care: A case study in New Mexico. *Southern Medical Journal* 1982;75:1110–7.
80. Zimitat C, Alexander H. *The 4SAT: A new assessment instrument for large classes in PBL curricula.* Paper presented at the 8th Ottawa International Conference on Medical Education. Philadelphia, PA: National Board of Medical Examiners, 1998.
81. Rezler AG. Self-assessment in problem-based groups. *Medical Teacher* 1989;11:151–6.
82. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Academic Medicine* 1991;66:762–9.
83. Kaufman DM, Hansell MM. Can non-expert PBL tutors predict their students' achievement? An exploratory study. *Academic Medicine* 1997;72:S16–8.
84. van der Vleuten CPM, Verwijnen M. A system for student assessment. In van der Vleuten CPM, Verwijnen M (Eds.), *Problem-based learning: Perspectives from the Maastricht approach*. Amsterdam: Thesis-Publisher, 1990.
85. Hebert R, Bravo G. Development and validation of an evaluation instrument for medical students in tutorials. *Academic Medicine* 1996;71:488–94.
86. Rolfe IE, Murphy LB, McPherson J. The interaction between clinical reasoning, group process, and problem-based learning: Assessment at the Newcastle Medical School. In Ostwald M, Kingsland A (Eds.), *Research and development in problem-based learning* (pp. 211–7). Sidney: Charles Sturt University Press, 1994.
87. Colliver JA, Williams RG. Technical issues: Test application. *Academic Medicine* 1993;68:454–60.
88. van der Vleuten CP, Van Luyk SJ, Beckers HJ. A written test as an alternative to performance testing. *Medical Education* 1989;23:97–107.
89. van der Vleuten CPM, Van Luyk SJ. Evaluating undergraduate training in medical skills. In ZM Nooman, HG Schmidt, ES Ezzat (Eds.), *Innovation in medical education: An evaluation of its present status* (pp. 404–21). New York: Springer, 1990.
90. Clarke R. The new medical school at Newcastle, New South Wales. *The Lancet* 1978;1:434–5.
91. Neufeld VR. Student assessment in medical education: A Canadian case study. *Assessment and Evaluation in Higher Education* 1982;7:203–15.
92. Leeder SR. The new pathway in general medical education at Harvard Medical School. *Medical Journal of Australia* 1991;155:740–3.
93. Des Marchais JE. From traditional to problem-based curriculum: How the switch was made at Sherbrooke, Canada. *The Lancet* 1991;338:234–7.
94. Areskog NH. The Linköping case: A transition from traditional to innovative medical school. *Medical Teacher* 1995;17:371–6.
95. Huber P, Perrier A, Balavoine J-F, Archinard M, Lefebvre D, Vu NV. Design and development of the new preclinical practice skills program at the University of Geneva. In A Scherpbier, C van der Vleuten, J Tethans (Eds.), *Advances in medical education* (pp. 679–81). Dordrecht, The Netherlands: Kluwer, 1997.
96. Hoftiezer V, Iatridis PG, Bankston PW, Vaughn S. Student evaluation in the regional center alternative pathway—a problem-based learning curriculum—compatible with a traditional grading system. In *Nineteenth Annual Session Innovations in Medical Education (IME) Exhibits* (p. 148). Boston: Association of American Medical Colleges, 1994.

Received 25 January 1999
Final revision received 31 March 1999