ASSESSMENT AND EVALUATION IN PROBLEM-BASED LEARNING

Robert Waters and Michael McCracken
Georgia Institute of Technology

While Problem-Based Learning (PBL) is frequently a culture shock to new students faced with its alien paradigm, one might say the same thing about the shock to educators who try to use this technique in their classes. Besides the fundamental challenge of creating a good problem, educators are faced with the task of deciding how to evaluate the technique's effectiveness and how to assess whether students have met the overall learning objectives for the course. It is our contention that traditional assessment techniques such as the familiar multiple-choice and true-false examinations do little to truly assess a student's understanding and far-transfer of the PBL learning experience. Likewise evaluating the success of PBL as compared to more traditional lecture-based classes requires more complex techniques. This paper assumes a basic understanding of PBL. Details of PBL in a medical setting are contained in Barrows[1] and in a computer science environment in Waters and McCracken[2].

PBL's student-centered focus and emphasis on self-directed learning create unique challenges for development of an effective assessment technique. In the medical school model, assessment occurs in the context of an individual PBL problem to be solved by the student [8]. While this method is most true to the PBL model, it is manpower intensive and not scaleable to large classes with only one or two instructors. On the other hand, traditional assessments in the form of objective tests scale well to large student populations, but are not true to the PBL model. In the field of assessment then, we seek to answer the question, "Is there a technique between the Barrows model and traditional assessments that can meet the needs of PBL and still be scaleable to large class sizes?"

In their paper on mathematics education the National Research Council (NRC) [3] relates a humorous quote that expresses the dilemma of educational assessments:

"You can't fatten a hog by weighing it."
--Anonymous Midwestern Farmer

While many might agree with this idea, it is also true that weighing a hog allows the farmer to detect when there are problems with the fattening process. Likewise student assessment seeks to provide a diagnostic tool to ensure students are progressing adequately towards achieving the desired learning goals. Assessment should not be solely a grade-assignment or ranking tool. All too often the learning process degenerates for students into striving to do well on the tests (assessments) so they will have a good grade, rather than focusing on the learning goals of the course.

The NRC contends that there are 3 guiding principles to assessment[3]:

- Content : assessment reflects what is most important for students to learn
- Learning: assessment enhances learning and supports instructional practice
- Equity: assessment supports every student's opportunity to learn.

In the first submission of this paper, the equity principle was summarily dismissed as being a bow to political correctness. The reviewers pointed out that this was a very misguided view and it failed to address the varied learning styles and needs of students. In that sense it is true that the equity principle is very important in selecting a pedagogical methodology. In fact, addressing diverse learning styles is a principle factor in the current interest in PBL by the educational community. This paper however is focused on assessment and evaluation, not in the actual pedagogical method used to teach. We assume the reader has already chosen PBL, and our challenge is to develop effective techniques for assessing students and evaluating the effectiveness of the educational intervention. In this sense, equity plays a much smaller role than other more important principles. We reject the notion that assessment standards for students should change based upon personal characteristics such as gender, culture or ethnic background. Assessment methods that are adaptable to PBL (which itself adapts to diverse learning styles) should -- by following the principles discussed later in this paper -- adapt to the needs of a diverse student population. For this reason, the equity principle will not be explicitly discussed further, but will remain an implicit part of the overall PBL methodology. On the other hand, the content and learning principles will be explicitly discussed in the following paragraphs.

PBL seeks to incorporate a multidisciplinary approach in the solution of problems. Students are given a problem, but not constrained on where they may look for answers. All too often, traditional assessment tools focus on isolated facts and techniques to the detriment of student's understanding of the larger integrated concepts involved. Narrow assessments allow students to pass through the system on rote memorization rather than true understanding. The concept principle emphasizes that assessments should never be trivialized for the convenience of assessment, but rather should emphasize problem solving, thinking and reasoning skills [3].

The learning principle emphasizes that assessments should continue the learning process and not be viewed as a disjoint activity. Too often there is a complete break in activities where teaching ends and then there is an assessment. This can encourage students to get by on their memorization skills and continue their passive participation in the educational process. Whatever assessment technique is used, it must be viewed by students as an active part of the learning process, not some ancillary activity. The idealistic goal of the learning principle is to make it indistinguishable to a visitor as to whether instruction or assessment is occurring in the classroom [3].

These principles certainly lay out laudable goals, but are there other principles that could be potentially applied to the assessment process? The National Science Foundation Education Standards propose several assessment standards that could be viewed as principles in designing effective PBL assessments [4]. After eliminating those that are reflected in the previous discussion, we are left with the idea that assessment should:

- Provide for authentic tasks.
- Provide data that are repeatable.
- Provide students the opportunity to evaluate and reflect on their own understanding
- Provide a means of reporting on student progress

Clearly, providing authentic tasks in assessment should be a key principle of any PBL assessment. An authentic task is one which is similar to a task performed in "real life." The uses and advantages of authentic tasks are discussed in situated learning [9]. If providing an authentic task as part of a good problem statement is a key to maximizing student learning during a PBL experience, why wouldn't it be key to the assessment experience also? If we want to blur the line between instruction and assessment, then one obvious solution would be to make the assessment be fairly similar to the instructional technique. As detailed by Anderson [5] however, authentic tasks alone are not the answer. There must be a close relationship between the assessment tasks and the desired learning outcomes of the class. Authenticity alone is not sufficient to ensure an adequate assessment.

Whatever assessment technique is used, the results should be repeatable. Using our hog analogy, if I weigh a hog in at a 100 kilos, then another identical hog should also weigh in at 100 kilos. This is an obvious goal, but the solution is not so obvious or easily attained. Other than some traditional trivial assessment techniques like multiple choice questions, most assessment tools , and especially those related to PBL, are fairly subjective. This makes it a challenge to create assessments where the results are repeatable over time.

Giving students the opportunity to evaluate and reflect on their own learning is a key element in PBL. The self-assessment phase of a PBL session allows the facilitator to help students with assessing their own performance in solving a problem. Self-assessment in PBL allows the students to compare their performance with the goals which they set for themselves before the problem started. It allows them to develop the skills to monitor their own learning outside the academic environment and helps them to move towards the elusive goal of becoming life-long learners. The ability of PBL to foster reflection in students is largely a function of the facilitator's ability to introduce opportunities for reflection into the PBL experience. If the facilitator fails to properly probe and question a student's conclusions, he does not foster an atmosphere where reflection can take place. An adequate assessment tool can help to overcome the deficiencies of any particular facilitator in fostering reflection.

Whether or not we consider "grades" as important, the requirement to report on student's progress will be around for quite some time. An effective assessment tool must also provide the ability to report on student progress in a fair and objective manner.

Given this list of principles in assessment, how do we then use them to design an effective assessment tool? The first step is to have a clear understanding of what the learning outcomes are for the course. This sounds trivial, but it often requires some thought to translate high level learning goals that might be published for a course into specific measurable knowledge elements that represent the key concepts we want a student to leave the course with. Fortunately, with PBL much of this work has already been accomplished since the same process is used in problem design.

There are two options for the next step. The first would be to prepare objective questions that test the student's comprehension of the tasks identified in the step above. The advantage of this is that it is fairly easy to do, and even easier to score. On the other hand it violates some of our previously outlined principles. First, it sacrifices depth for ease of scoring. Secondly, and probably most importantly, it violates the very principles we espouse if we believe in PBL as a learning technique. We ask the students to conduct problem-solving through analysis of the facts, refinement of hypotheses and examination of learning issues as a way to effectively solve ill-structured problems. Then during assessment, we ask fragmented questions in a traditional manner. Students are often puzzled and feel betrayed by an apparent shift in instructor techniques.

The second option should sound familiar for any PBL enthusiast -- create a problem statement the solution of which requires the student to demonstrate the desired depth of understanding of the learning outcomes. This is similar to the Barrows model previously referred to. The problem can be similar to the ones used in class which would assess near transfer, or they could be apparently very different on the surface which would assess far transfer. Of course, to be feasible for large group PBL, the solution needs to be achievable within the assessment time available. This creates a challenge for the problem designer that is potentially even more difficult than the design of a normal PBL problem.

As outlined in our previous paper [2], we use PBL to teach a senior undergraduate course in Management Information Systems (MIS) Design. For fall quarter 1996, the class was composed of 5 groups of 6 students each, with one instructor and 2 teaching assistants for the class. We began the course by having the students personalize the desired learning outcomes to develop an individual educational prescription as described in Barrows[1]. We then had the students do self-assessments at the end of each phase. The self-assessments followed the general outline of Woods[6]. During the self-assessment phase, each student was encouraged to discuss the quality of learning resources used and how well they had progressed towards meeting their individual learning objectives.

The problem for the course was to design an interactive auction system for a specific auction establishment. Students interviewed customers and developed the initial requirements from which they would develop their design and subsequent implementation. At each stage of the problem, they turned in the same artifacts that would have been produced in a regular software development project. At the end if each of the phases, the students also did both self-assessment and group assessments for that particular phase. By assessing group performance at each phase, the instructor could detect group problems and try to apply corrective action. The assessment of the artifact provided an indication of how well the students did at each phase and whether any major concepts were not learned. Perhaps the most interesting observation is the student reaction to the first artifact, the requirements document.

The student's performance on the requirements document has always been consistently poor across all the occurrences of the course and within groups in the same course occurrence. There are several reasons why we believe this occurs. First, as any software engineer will relate, the requirements phase of a project is the most difficult. It is here where a "blank" sheet must be transformed into a specification of exactly what the customer desires. Secondly, this is the most ill-structured portion of the problem. There is little information to begin with, there are multiple techniques that may potentially be used, and there are multiple paths to several acceptable solutions. (It is this very ill-structuredness that first led us to research the PBL process for software engineering education in the first place.) Thirdly, students in software engineering up this point have typically had little experience in the requirements gathering process. They have received problems in previous courses that have had most of the initial specification completed and thus they had no need to actively elicit and present detailed requirements information. Because of this, students usually take too narrow a view of the problem and focus too soon on potential solutions rather than understanding the many implications of the problem requirements. Finally, while the difficult requirements process is occurring, the students are also making their transition into the PBL process. Up to this point few -- if any -- students have been exposed to PBL. Most are still comfortable only in the traditional teacher-centered classroom.

Addressing these problems are somewhat difficult. Clearly not much can be done to alleviate the difficulty inherent in the requirements process itself. While significant research is being conducted to find effective and efficient means of gathering requirements, no general technique has yet been found. The only areas available for assistance are in helping students structure their knowledge and easing their transition into the PBL process. This quarter, we conducted a PBL demonstration, used a textbook about PBL in the course and tried to closely monitor the groups progress. Still as in years past, students still did poorly in the requirements phase. Whether student difficulty is dominated by the difficulty of the requirements process, or whether there is something we are lacking in our PBL implementation is still an open research question.

Student performance improves dramatically during the design and implementation phases. Intuitively, we might expect this to happen for several reasons. First, students have done design and implementation in previous courses and are able to transfer this knowledge into the MIS design realm. Secondly, after preliminary design is completed, the problem has become structured enough so that the students can generally complete the solution without major complications. This second observation has identified another open research question which we are pursuing -- whether PBL is the appropriate paradigm to follow after the students complete their preliminary design phase. Even in the medical model of PBL, after the student's learn their basic sciences, they move to the clinical phase of instruction. Here the self-directed learning skills developed in the earlier phases are used by the medical students, but not in the explicit PBL group paradigm. This leads us to hypothesize that for the later stages of the problem solution, some other paradigm than traditional PBL might be more apropos.

The use of software engineering artifacts as assessment tools meets most of the assessment principles discussed earlier. They satisfy the content principle because they are the documents that explicitly detail the problem solution and allow the practical expression of the course learning objectives. Writing and presenting these documents is integrated into the PBL process so that their preparation and the basic learning activity are indistinguishable. Certainly, these are authentic activities since they directly correspond to real-world software development artifacts. The detailed comments and discussion of the deficiencies in the various artifacts provide a good means of student reflection and self-assessment. After the requirements phase in particular, students generally reflect on their performance and gain insight into the true difficulties and challenges of gathering and presenting a good specification. Grading the artifacts using a detailed criteria provides a mechanism to report on student progress. The only criteria for assessment that is not totally met is repeatability. As with any essay or report grading, there is variability in the basic grading. We use a detailed grading criteria for each artifact so that students understand what should be included, and so we can introduce consistent grading criteria. However, there is always variance in this type of grading that cannot be totally eliminated. On the whole however, we feel the use of the graded artifacts is an excellent means of assessment in PBL.

The student-produced artifacts do differ from real-world software artifacts in one important area. The students must discuss their reasoning processes in the solution of the problem for the particular phase of development they are preparing the document for. For instance, in the requirements phase, they must justify the methods they chose both to elicit requirements and to represent those requirements. This helps the students reflect on the knowledge they have gained during this phase and provides an indication of whether there are any misunderstandings of the course material. It also helps to eliminate the possibility that the students simply copied information from a similar problem in a reference text without understanding what they were actually doing. This concept is similar to the popular "learning essay" technique, but differs in that we require the students to integrate reflective activity into the production of the real document rather than as a separate one. We feel this integration is more in keeping with our previously presented assessment principles. Integrating reflective activity into the artifact production also helps us to overcome the problem of a weak facilitator who might not elict reflective activity from the students.

We have also used more traditional assessment methods such as the infamous midterm and final examinations. For these examinations we chose to present the students with small problems that they could solve independently within the allotted time. The examinations were open-book, and were designed to assess overall student grasp of important concepts in the course. For instance, at midterm the students had completed the requirements phase and were just beginning to transition to preliminary design. The midterm consisted of four questions which covered the major concepts that students should have understood up to this point in the PBL process. The problems were stated as authenticly as possible so students could see the application of the knowledge required to actual software development practice. Like the artifacts already discussed, the examinations generally met our assessment criteria. They suffer from the same repeatability problem and also have the added disadvantage of not being as authentic a task as preparing the actual software artifacts.

The final examination consisted of a take-home examination which comprehensively covered all learning outcomes for the course. The format was generally identical to the midterm examination. One question was specifically designed to measure far transfer and one was designed to generate student reflection and overall self-assessment. The remaining questions were small problems oriented towards assessing student grasp of key concepts.

The other side of determining how well PBL is working is evaluation. The National Science Foundation has published a very useful booklet which every practitioner contemplating an evaluation project should possess [7]. NSF outlines three types of evaluation: planning, formative and summative. Planning evaluations, while important are not applicable to our established program so they are not further discussed. Formative and summative evaluations are applicable and are best understood by using a quote from [7]:

"When the cook tastes the soup, that's formative,
When the guests taste the soup, that's summative."

--Bob Stake

We wanted to design an evaluation of the PBL process that allowed us to determine whether PBL was more appropriate to the training of software engineers than other more traditional methods. Since technology changes so rapidly in the computer field, PBL's self-directed learning component is a key feature of interest. Unfortunately, the best evaluation would be to examine software engineers several years after completion of PBL and assess their ability to assimilate new information as opposed to those who received a traditional lecture-based education.

Since this would require a long term study, we wanted to focus on methods more appropriate for a shorter term evaluation method. Questionnaires were chosen to gather qualitative information from students as part of the formative evaluation. We wanted to capture student opinions and ideas on improving the class.

The summative evaluation consisted of course observations by the TA's and instructor and comparison of pre-test scores to post-test scores. The TA's met regularly with the PBL groups and were in a position to note problems in the process. These problems and general observations about the process were captured through informal interviews with group members. Additionally, the students were administered an examination during the first class period which covered all the major learning outcomes for the course.

This pre-test examination was scored and the grades compared to the final examination scores to provide an indication of overall student learning and transfer in the course.

One real difficulty in PBL evaluation as compared to evaluating other more traditional classroom interventions, is in the nature of the group dynamics that occurs during the learning experience. This produces variation in results not only between classes, but between individual groups within the same class. This is why regular out-of-class meetings between the TA's and the individual student groups was so important to the overall evaluation of the class.

We are in the process of analyzing out evaluation data to determine potential improvements in the course structure. We also hope to find answers to the open research questions previously described.

An effective assessment and evaluation program can insure that students are deriving the maximum benefits from PBL and that the PBL process itself is being conducted effectively for the given environment. This paper presented several general principles for assessment of PBL classes, and then reviewed the techniques actually used against these principles. In our experience thus far the requirement to have authentic tasks conflicts with the requirement for assessments to be repeatable. We believe this is primarily caused by the fact that the authentic tasks are themselves ill-structured and thus difficult to assess completely objectively.

There are also many areas in PBL and assessment that are still open issues. The usefulness of PBL after the problem becomes structured and the best way to improve student performance early in the course are two of these.

REFERENCES

[1] Barrows, H. and Tamblyn, RM, *Problem-Based Learning: An Approach to Medical Education.* New York: Springer Publishing Company, 1980.

[2] Waters, R. and McCracken, M., *Problem-Based Learning in Computer Science*, 5th Annual Conference on Problem-Based Learning, June 96.

[3] Mathematical Sciences Education Board, *Measuring What Counts, A Conceptual Guide for Mathematics Assessment*, National Academy Press, 1993.

[4] National Science Education Standards, *Assessments in Science Education*, National Academy Press, 1994.

[5] Anderson, J.R., Reder, L.M. and Simon, H. A., "Situated Learning and Education", *Educational Researcher*, Vol 25, No 4, May 96, pp5-11.

[6] Woods, Donald, *Problem-based Learning: How to Gain the Most from PBL*, McMaster University: 1994.

[7] Stevens, F. et al, *User-Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering and Technology Education*, NSF: 1996.

[8] Barrows, H., Presentation at Georgia Institute of Technology PBL Workshop, Dec 1995.

[9] Cognition and Technology Group, "Anchored Instruction and Situated Cognition Revisited", *Educational Technology*, March 93.