High Educ (2011) 62:1–16
DOI 10.1007/s10734-010-9361-3

Student and tutor perceptions on attributes of effective problems in problem-based learning

Nachamma Sockalingam · Jerome Rotgans · Henk G. Schmidt

Published online: 12 August 2010

Abstract This study aimed to identify the attributes that students and tutors associated with effective PBL problems, and assess the extent to which these attributes related to the actual effectiveness of problems. To this end, students and tutors in focus groups were asked to discuss about possible attributes of effective problems. The same participants were then asked to individually and independently judge eight sample problems they had worked with. Text analysis of the focus group discussion transcripts identified eleven problem attributes. Participants' judgments of the sample problems were then frequency-scored on the eleven problem attributes. Relating the participants' judgments with the entire student cohort's grades yielded high and significant correlations, suggesting that the eleven problem attributes reflect aspects of problem effectiveness.

Keywords Problem-based learning · Attributes of problems · Problem effectiveness · Problem evaluation

Introduction

Problem-based learning (PBL) is an approach to learning and instruction that has the following characteristics: (1) the use of problems as the starting point for learning, (2) small-group collaboration, and (3) flexible guidance of a tutor. Since problems steer the learning in such curriculum, (4) the number of lectures are limited. The latter is in line with

Part of this paper was presented at the 2008 annual meeting of the American Educational Research Association, New York.

N. Sockalingam (✉)
Republic Polytechnic, 9 Woodlands Avenue 9, 738964 Singapore, Singapore
e-mail: nachamma_s@rp.sg

J. Rotgans
Republic Polytechnic, Singapore, Singapore

H. G. Schmidt
Erasmus University Rotterdam, Rotterdam, The Netherlands

the idea that (5) learning is to be student-initiated and that (6) ample time for self-study should be available (Hmelo-Silver 2004; Schmidt 1993). Since all learning in PBL originates from the confrontation with a problem, its attributes are potentially of importance. Well-designed problems may, in principle, lead to better learning. Indeed, existing studies demonstrated the impact of problem quality on students' learning. For instance, Gijselaers and Schmidt (1990) investigated how the quality of problems related to the other aspects of PBL such as: students' prior knowledge, tutor performance, group functioning, time spent on individual study, achievement, and interest. Results showed that compared with students' prior knowledge and tutor performance, the quality of problems had the most influence on group functioning and time spent on individual study, and through these on interest in subject matter and academic achievement (See also Schmidt and Gijselaers 1990; van Berkel and Schmidt 2000). The findings imply that a high-quality problem is likely to produce a stronger positive impact on the learning process and outcomes than tutor performance and students' prior knowledge.

Problems are typically a set of descriptions of a phenomena or situations in need of explanations (Schmidt 1983), often presented in textual format, sometimes with illustrations, pictures, videos, and simulations. They are also sometimes known as "cases", "triggers" and "tasks". Problems are purported to engage the students, rekindle their prior knowledge, spark discussions, encourage collaborative work, promote self-directed learning skills and lead to acquisition of relevant content knowledge in the course of tackling the problem (Barrows and Tamblyn 1980; Hmelo-Silver 2004). When a problem is presented to students to initiate the learning process, the students confront the problem using their own prior knowledge and knowledge offered by their teammates. Issues emerging from the group discussion that demand further exploration are used as guidelines by the students for their self-directed learning activities. Following a period of self-study, they reconvene to discuss, share information, and synthesize answers to their queries as a team, integrating their new knowledge in the context of the problem. Finally, they report their findings to the class, and reflect on their learning (Schmidt 1983). Overall, the learning process in PBL is self-directed by the students and is more problem-directed than teacher-directed.

The role of teachers in PBL is considerably different from the role of teachers in a conventional curriculum, not only because they have a different name: tutors. While it is true that PBL tutors are actively engaged in the students' learning process like the teachers in conventional curriculum, the critical difference is that the PBL tutors do not directly transmit/teach the content knowledge to students; instead they facilitate the students' learning process by observing the students, stimulating discussion amongst team members, raising thought-provoking questions, encouraging collaborative work, and providing feedback at appropriate instances to the students (Das et al. 2002; Maudsley 1999). In PBL, it is the students who take the responsibility to synthesize the content knowledge through self-directed learning and group discussion which is in turn determined by the nature of problems.

Transition in the roles of tutors and students in PBL, and the intended purpose of problems emphasize the importance of problems in the learning process. This means that designing effective problems in PBL is quintessential to help students learn better. So, how do we design effective problems? In other words, what are the attributes of effective problems? Identifying the attributes of effective problems would not only shed light on how to design effective problems, but it would also help us in evaluating the effectiveness of problems. Furthermore, it could improve our current understanding the process by which the problem quality influences students' learning. To answer these questions and understand what is known about problem attributes, we carried out a literature review.

The literature review revealed that traditionally, guidelines derived from cognitive theories and experiential knowledge served as the basis for designing problems. Notably, Shaw's five problem attributes (1976), Dolmans' seven principles of case design (1997), and Hung's (2006) conceptual framework for designing PBL problems provided theoretical dimensions of problems. Surprisingly, there were limited empirical studies describing problem attributes. These empirical studies could be broadly categorized into two groups: those based on students' perceptions and those based on tutors' perceptions.

Let us start with the studies that explored problem attributes based on students' perceptions. Dolmans et al. (1995) investigated the effectiveness of problems by comparing the learning issues generated by students with those intended by faculty for twelve problems. The idea behind the comparison was that an effective problem will lead the students to the intended learning issues which could be measured by assessing the match between the student-generated and faculty-intended learning issues. The results suggested that students identified an average of 64% of the intended learning issues across the twelve problems, the gaps in identifying the learning issues were attributed to the complexity and unfamiliarity levels of the problems.

To understand what complexity and structuredness of PBL problems mean to students, Jacobs et al. (2003) utilized a questionnaire. They defined complexity of a problem as the interrelationship and stability of a number of characteristics that play a role in challenging the students to learn. Structuredness of a problem was defined as the application of a limited number of organized rules to tackle problems in a direct and predictable way. Based on this definition, well-structured problems have one clearly defined solution and ill-structured problems have many possible solutions. Their results showed that while students could differentiate between simple and well-structured problems, they had difficulty discerning between ill-structured and complex problems. Students also considered problem structuredness to be more important in determining the problem quality.

Looking at a different aspect of PBL problems, Soppe et al. (2005) investigated the influence of problem familiarity on learning process and achievement. In their experimental study, students were randomly presented with either a familiar or unfamiliar version of a problem. Students' self-report on the problem attribute familiarity and various other indicators of their learning such as self-study time, number and quality of explanations generated were used as measures in the study. Their result suggested that although familiarity of the problem influenced interest in working on the problem, there was no significant influence on academic achievement.

A common theme amongst studies based on students' perceptions is that they tend to focus on specific few problem attributes. On the other hand, studies based on tutors' perceptions explored a wider array of problem attributes. For instance, Kim et al. (2006) reviewed one hundred studies from various disciplines with the objective to classify a wider spectrum of problem attributes in PBL. They delineated a total of five problem attributes; they are that the problem should be relevant, realistic, engaging, challenging and instructional (build upon prior knowledge). Taking a different approach, Des Marchais (1999) used a Delphi study to gather six experts' opinions on what makes a good problem. He identified a total of nine attributes. These attributes were that good problems should: (1) stimulate thinking, analysis and reasoning, (2) assure self-directed learning, (3) enable use of prior knowledge, (4) be set in a realistic context, (5) lead to the formulation of appropriate learning goals, (6) arouse curiosity, (7) include topics related to public health (the study was conducted in a medical context), (8) assure contextual breadth, and

(9) choose an appropriate vocabulary. Of these, problem stimulating thinking, analysis, and reasoning and lead to self-directed learning were considered by the experts to be the two most important attributes.

Both these studies identified similar attributes of problems, adding support to the reliability of the findings. Furthermore, Des Marchais (1999) study was the first to identify a wider array of problem attributes. However, this study had two limitations. One, the expert responses were generalized opinions about PBL problems; these opinions were not based on specific problems judged. While this approach is perhaps useful in formulating a general perspective on problems, it does not illuminate the concrete experience of a particular problem. To achieve this, one possibility is to present participants with concrete examples of problems to judge. Two, only expert judgments were considered in their study. As students are the end-users of the problem, we felt that investigation of students' opinions and comparison of the students' and tutors' opinions about the quality of the problems will be useful.

In summary, most of the existing literature on problem attributes do not incorporate both students' and tutors' perceptions. In addition, asking participants to mention desirable attributes of problems in general may yield different answers from asking them to mention attributes of specific problems. Finally, with the exception of Soppe et al. (2005), most studies did not try to relate problem attributes directly to academic achievement. This warrants a need for additional studies on problem attributes. To this end, the present study aimed to answer the following questions: Which problem attributes do students and tutors generally consider as contributive to the overall effectiveness of problems in PBL? Do the students' and tutors' perceptions of problem attributes hold across a set of problems? To what extent do students and tutors agree in their judgments of the overall effectiveness of these problems? Does the evaluation of problem effectiveness, based on the identified attributes, reflect itself in the students' academic achievement?

To address the above questions, we conducted focus group interviews with a total of eleven students and five tutors in two phases. The first phase was a group discussion in which we sought the students' and tutors' generalized opinions about attributes of effective PBL problems. In the second phase, we asked the same students' and tutors' to respond individually on eight problems they have worked with.

Method

Participants

Eleven first-year polytechnic students and five tutors participated in the study. Both the students and tutors were randomly selected amongst those participating in a science module. The tutors taking part in the study had an average tutoring experience of 1 year and 7 months. In addition, we used student achievement data from the entire cohort of 2566 first-year students taking the science module.

Educational context

This study was conducted during the second semester of academic year 2006/2007, at Republic Polytechnic, Singapore. The polytechnic has adapted problem-based learning as its instructional method and has implemented it in a "one day, one problem" approach (Alwis and O'Grady 2002). This approach requires students to work on one problem per day. Each day, students spend their time on three meetings, with a self-study period between the meetings. In a typical class size of twenty-five, students are grouped in teams of five and are guided by one tutor. The students are presented with the problem in the first meeting and encouraged by the tutor to discuss what they know, do not know, and need to find out; in other words students define their own learning issues. The learning issues generated then serve as a basis for further exploration during the subsequent self-study period. During this first self-study period students search for relevant resources, read the resources, and exchange ideas with their teammates. Following this, the students and the tutor reconvene at a second meeting to discuss the progress. The second meeting provides the tutor with an opportunity to gauge the students' engagement and progress through discussion and observation. Subsequently, a second and longer self-study period provides students with the opportunity to explore the topic in more detail to fill gaps in their understanding, to compile the information collated and to prepare for a presentation during the third and last meeting. During this third meeting, the students present their findings to the class, answer questions, and clarify doubts. The day ends with an opportunity to reflect on their learning by means of keeping an electronic journal.

Materials

Problems

Eight problems, familiar to both students and tutors, were used in the study. The science module is structured in such a way that it provides an introduction to foundational, interdisciplinary scientific principles and applications. The module comprised of sixteen problems in total, covering various topics like Cells, Recombinant DNA Technology, Energy, Electricity, Atomic Structure, and Structure of Organic Compounds. Of the sixteen problems, the first eight problems were used in this study. These eight problems are referred as P1 to P8. Five of the eight problems, P1, P2, P5, P6, and P7 were biology-based whilst three problems, P3, P4, and P8, were physics-based. The biology-based problems focused on Structure and Function of Biological Materials as well as Genetic Expression. The physics-based problems focused on Heat Transfer and Properties of Light. All eight problems were in text format with P4 being the longest problem at two and a half A4 pages length. The other problems were shorter than one page. Additional features of the problems were that P3, P4, and P 8 contained either pictures or diagrams, whilst P5 included an excerpt from the poem "Heredity" by Thomas Hardy. A copy of the biology-based problem P2 and physics-based problem P4 is attached in the appendix A for reference.

Students' achievement measure

Students' academic achievement, referred to as the *daily-grade*, was recorded by the tutors after every problem. The daily-grade is based on competencies demonstrated by the students during the course of the day, such as participation in discussions, teamwork, time and resource management, ability to collate relevant information, demonstration of reasoning skills, indication of critical thinking, and evidence of understanding. The students were graded on a 5-point performance scale: 0 (*fail*), 1 (*conditional pass*), 2 (*acceptable*), 3 (*good*), and 4 (*excellent*). For each student, one daily grade score was recorded for each problem. It has been shown elsewhere that the daily-grade demonstrated high levels of reliability (Chai and Schmidt 2007). Their findings were based on 1,059 student observations by 230 tutors, which resulted in generalizability coefficients ranging from .55 to .94

(average = .83). In addition, this measure correlated .47 with the results of a written achievement test. These values are indicative of a high reliability and good predictive validity of this measure.

Procedure

The eleven students who participated in this study were randomly grouped into one team of three and two teams of four, whilst the five tutors were grouped into one team of two and one team of three. Thus, tutors and students were not in the same group. The focus group interviews were conducted in two sequential phases. The first phase involved group discussions whilst the second phase demanded independent responses from the individual participants. Each phase took 45 min on average and participants completed both phases in a single stretch. In the first phase, the participants in their respective groups were asked to discuss "What is an *overall* effective problem to you, based on your experience?" The focus group discussions were audio-recorded and transcribed for further analysis.

Following the discussion, the same participants were presented with eight familiar problems that they have worked on previously in the academic course and asked to individually write down responses to the question "what are the positive and negative aspects for each problem?" They were also informed that *positive aspects* of the problem refer to the problem attributes deemed as contributive to the overall effectiveness of the problem whilst *negative aspects* refer to attributes deemed as contributive to the overall ineffectiveness of the problems. The rationale for asking the participants to give the reasons for the effectiveness of the problems (instead of asking the participants to rank a given list of problem attributes) is to gather more insights into why participants may consider a certain problem to be effective.

To conduct the second phase, the eight problems were displayed separately on eight designated tables. Writing materials such as note-pads, pens and a folder to post the completed written reports were made available in the designated tables. The participants were instructed to proceed to any unoccupied tables to read the problem, record the positive and negative aspects of the problem on the given notepad, and post the completed note in the folder placed at each table before moving onto the next table until they had visited all eight tables. The setup was such that the participants had no opportunity to talk or read the notes of the other participants. At the end of this phase, we collected the notes from the five teams for further analysis.

Analysis

The data analysis in this study was designed to be sequential in that interpretation of the data from the second phase was dependent on the results from the first phase. The transcript data from the first phase of focus group discussion were analyzed using TextSTAT software obtained from the web link, http://www.niederlandistik.fu-berlin.de/textstat/ (Huning 2007). TextSTAT is a simple concordance program that is designed to count words in the input text. The program uses text files in ASCII/ANSI/HTML/Microsoft Office formats and generates a frequency-list of words in Microsoft Office formats. From the list generated, appropriate evaluative words associated with the various qualitative aspects of problems were manually identified and categorized as various problem attributes. The assumption was that the more often an attribute-associated word was mentioned, the more important the attribute was for the respondent.

The following example illustrates how the text-analysis was conducted. A transcript excerpt from a tutors' response to the question used in the first phase is "*Every problem has to have learning objectives. Students should be challenged to look into solving the problem. It should be motivating enough so that the students feel like doing it enthusiastically". When this excerpt was used as an input file, the textSTAT software generated a frequency list of 26 words. From the list generated, words such as *solving*, *doing*, *like*, *motivating*, *challenged*, *enthusiastically*, *learning*, and *objectives* were manually identified to be connoting problem attributes. These words were then classified based on semantic similarity into three problem attributes: *problem interestingness* (like, motivating, challenged, enthusiastically), *problem promoting self-directed learning* (solving, doing), and *problem leading to learning goals* (learning, objectives).

To analyze the data from the second phase of the study, the newly identified problem attributes were used as the criteria to frequency-score the participants' individual responses about the effectiveness of the eight sample problems. The following example using three student responses about problem P4 illustrates how the frequency-scoring was carried out. One student reported P4 to be positive as it was "interesting and makes people question". A second student reported P4 to be "interesting, because it makes us think about how it happens", and a third student reported P4 to be "story-like question, interesting". In this case, the frequency score was computed as one count for *problem format* (story like), two counts for *problem stimulates critical thinking* (makes people question, makes us think) and three counts for *interestingness of problem* (interesting). In a similar fashion, the eleven students' responses on the positive aspects were scored for problem P4 and the other problems. The summative score obtained is referred as the *observed positive student score* for the respective problem. Following this methodology, the *observed negative student score, observed positive tutor score*, and *observed negative tutor score* were computed for each of the eight problems. Subsequently, the negative score for each problem were subtracted from the positive score for both groups to obtain the *observed overall student score* and *observed overall tutor score* for each of the eight problems.

To investigate if the eight problems in fact differed in terms of effectiveness, a one-way Chi-square test was carried out. To do this, absolute values of the *observed overall student score* for each of the eight problems were compared with the *expected overall student score*, assuming the null hypothesis that there was no difference between the problems in terms of effectiveness. This was not repeated for the tutors due to the low frequency of some of the responses.

Following that, a Chi-square test of independence was used to investigate if students and tutors differed in their judgments. This was done, first for the positive and then the negative aspects, by comparing students' and tutors' *observed* frequency-scores with the *expected* frequency-scores. The null hypothesis for both the comparisons was that there was no significant difference between the students and tutors in their judgments of the eight problems in terms of the positive and negative aspects. For each problem, we compared the sum of all the scores of the eleven problem attributes rather than the scores of the individual problem attributes as the responses on some attributes were insufficient for Chi-square analysis.

To investigate if the identified problem attributes influenced the effectiveness of problems, the tutors' judgments of problems, represented as the *observed overall tutor score*, were correlated with the average grades obtained by the entire cohort of 2566 students taking the science module for the respective problems using Pearson's correlation coefficient.

Results

Text analysis of the discussion transcripts revealed that the students and tutors associated a total of eleven problem attributes with effective PBL problems in general. The identified attributes ranked as per the frequency of the connoting words are presented in Table 1.

Both students and tutors agreed that an effective problem should first and foremost lead to the appropriate learning goals. However, as Table 1 shows, there were some differences between students and tutors in ranking the remaining attributes.

Next, a cursory scan of the student and tutor responses on the positive aspects (effectiveness) and negative aspects (ineffectiveness) of the eight sample problems did not reveal any other additional problem attributes. In other words, the eleven general problem attributes were held in considering the effectiveness of problems at the micro-level of specific problems. Hence we used these attributes as criteria to frequency-score students' and tutors' responses with regard to the effectiveness of eight sample problems. The computed *overall* student and tutor frequency-scores for the eight problems are shown in Table 2.

On the whole, P4 was considered to be the least effective problem while P2 was considered to be most effective problem by both the students and tutors. A copy of the problem P2 and problem P4 is attached in the appendix. A sample of students' responses about the negative aspects of P4 was that:

> A good example of long winded problem statement (3 pages long), one would have to read over and over again to make sure nothing is lost; Too long a problem statement, we lose track/forget what we have read half way through; Story too long–testing our patience, not everyone will read it, as our laptop is just in front of us and many temptations.

Tutors responses echoed the students' views. They mentioned that the problem was:

> Too long, examples were not so practical, students turned off before it started; A bit lengthy, probably unavoidable as well due to the complex nature of underlying concepts; Too long, scared off the students, wasting a lot of time to read repeatedly.

Table 1 Key attributes of effective problems according to students and tutors

Students* Effective problems should…	Tutors* Effective problems should…
Lead to appropriate learning goals (37.4%)	Lead to appropriate learning goals (25.1%)
Promote self-directed learning (22.1%)	Promote self-directed learning (24.6%)
Stimulate critical thinking (14.5%)	Trigger interest (16.0%)
Promote teamwork (10.7%)	Be of suitable format (8.2%)
Trigger interest (4.6%)	Stimulate critical thinking (5.9%)
Be of suitable format (3.8%)	Relate to prior knowledge (5.8%)
Be of suitable clarity (2.3%)	Enable application/be of relevance (3.8%)
Stimulate elaboration (2.3%)	Promote teamwork (3.8%)
Enable application/be of relevance (1.2%)	Stimulate elaboration (3.1%)
Relate to prior knowledge (0.8%)	Be of suitable clarity (2.0%)
Be of appropriate difficulty (0.4%)	Be of appropriate difficulty (1.8%)

* Problem attributes were ranked according to frequency of connoting words

Table 2 Observed student and tutor frequency-scores of positive and negative aspects of eight sample problems

Responses	P1	P2	P3	P4	P5	P6	P7	P8
By students								
Positive	23	25	14	13	18	19	11	19
Negative	5	11	10	24	7	6	7	6
By tutors								
Positive	5	11	6	3	7	7	8	9
Negative	8	5	6	13	6	4	2	5

Yet, students considered P4 to be effective in the sense that it made them think and reason, thereby stimulating critical thinking. They mentioned that the problem was "interesting, makes people question". However, some tutors commented "Nil" to positive aspects of P4 and tried to reason that "Most likely this is one of the simplest ways a student may be introduced to the quantum nature of light". As for the most effective problem, P2, some of the students felt that it was:

> Short and easy to understand, problem related to us (our human body), make us want to go find out more about our body; Quite basic, was able to relate to what I have learnt in secondary school (prior knowledge); It is straight to the point to what we are supposed to find out; Short direct, gives clues from the problem statement, tells us what we have to do.

Similarly, some of the tutors felt that P2 was "Practical, examples capture the interest, it is good that it specifies the learning objective; Problem statement is short and to the point, embedded concepts are very intriguing; Short and nice".

> As for the negative aspect, some of the students' comments are as follows:
> No pictures/figures, perhaps we can give analogies to let students understand better; first paragraph too complicated, not easy to comprehend if we have bad command of language, second paragraph tends to confuse student/tutor. Putting together makes it complicated and time consuming; the terminologies they use is quite difficult to understand.

Tutors did not have much negative comments. Some of their responses on negative aspects of the problem were "Nil; Instead of asking the students to examine, how about asking them a question on how different cells act in the same yet in such diversified ways in order to self sustain".

A comparison of the absolute values of the *observed overall student score* with the *expected overall student score* for each of the eight problems using Chi-square analysis showed that the students found the eight problems to be significantly different from each other, χ^2 (7, $N = 88$) = 14.91, $p = .04$.

Next, a comparison of the *observed* student and tutor frequency-scores with the *expected* student and tutor frequency-scores for the positive aspects of the problems indicated no significant differences, χ^2 (7, $N = 198$) = 4.34, $p = .74$. Likewise, comparison between the student and tutor responses on the negative aspects of the problems indicated no significant differences, χ^2 (7, $N = 125$) = 4.96, $p = .67$. In sum, the results suggest that there was no significant difference between the students and tutors in their judgments about the effectiveness of the eight problems.

Finally, a correlation of the tutor judgments represented by the *observed overall tutor score* for the eight problems with the student grades obtained by the entire cohort of 2,566 showed a high, significant and positive correlation, with an r value of .75, $p < .05$. Likewise, a correlation measure of the *observed overall student score* and the average daily-grade showed a high, significant and positive correlation, with an r value of .82, $p < .05$.

Discussion

The purpose of this study was first to explore both students' and tutors' generalized perceptions of attributes associated with effective PBL problems. Second, verify if these generalized perceptions are held in judging the effectiveness of specific problems. Third, to examine the extent to which the students and tutors agree in their judgments of the overall effectiveness of the sample problems, and fourth assess the extent to which these attributes actually relate to the problem effectiveness. Results suggest that both the students and tutors associated a total of eleven attributes with the effectiveness of PBL problems in general, which were also considered by the students and tutors in judging specific problems. There was no difference between the students and tutors in their judgments about the overall effectiveness of the eight sample problems, and the student and tutor judgments about the effectiveness of the eight sample problems correlated significantly and highly with the student grades. Overall, the identified eleven problem attributes, derived from both student and tutor opinions, turned out to be related to student learning.

In answering the first question, we generated eleven attributes of effective problems based on focus group discussion with both students and tutors. To see if the eleven attributes measured the same aspects mentioned in the other studies, the attributes generated in this study were compared with those cited in the earlier mentioned literature. The eleven identified attributes of effective problems largely covered the attributes mentioned by the various other studies, including the nine attributes mentioned in Des Marchais' study (1999). We found three attributes to be common amongst the various studies. They were: (1) a problem should lead to formulation of appropriate learning goals, (2) a problem should relate to the students' prior knowledge, and (3) a problem should be interesting. It is not clear why some studies generated certain attributes but not others. Nevertheless, we infer that the commonality of the three attributes regardless of the differences in the various studies imply importance of these attributes in designing problems.

In addition, this study identifies a unique attribute not mentioned in the other studies—*problem format*. The problem format refers to the physical structure of the problem, and includes features such as the length of the text, use of appropriate pictures, illustrations, videos, and simulations in the problem. Although not much is mentioned about problem format in PBL literature, research on instructional design suggests that the format of instructional material may influence the learning efficacy of a learning environment as a result of the cognitive load required (Hoffler and Leutner 2007). Cognitive load refers to the load exerted in learning in terms of working memory and it can be classified into three types, namely, intrinsic, extraneous, and germane. The cognitive load of concern here is the extraneous cognitive load which refers to the working memory load that learners experience as they interact with the instructional material. Even though several studies have been carried out on the influence of extrinsic cognitive load on learning (e-learning and multi-media learning), there is a need for studies on the influence of extrinsic cognitive load in constructivist environment (Morena and Park 2010) such as PBL. As such, an investigation of the influence of problem format on students' learning process and outcomes in PBL may be worth exploring.

Next, when we compared whether the students and tutors differed in their generalized perceptions of PBL problems, we found that the students and tutors alike emphasized that an effective problem should lead to appropriate learning goals. This attribute is also considered by Dolmans et al. (1995) and Mpofu et al. (1997) as an important attribute of effective problems.

The implications of these findings for designing PBL problems are that problem designers need to carefully consider the intended learning goals of a problem and formulate the problem such that it is clear and guiding towards the learning goals. To make the problem clear, problem designers could provide pictures, keywords, examples or analogies which relate to students' prior knowledge. This would in turn help the students towards the learning goals of the problem. As problem format seems to have an impact on the effectiveness of problems, we recommend that problem designers try different formats before selecting one. The choice of problem formats could be based on an analysis of the learners' needs, learning style, and the suitability of the format with the intended learning objectives. For instance, in the case of the problem P2 used in this study, instead of the textual description given, the problem could utilize an analogy, comparing a house with a cell and the various rooms in the house with different compartments in the cell. Alternatively, one could present pictures of a cell depicting different compartments requiring the students to examine how the cells are structured to self-manage their systems and functionalities. Utilizing a suitable format is likely to help the students understand better and engage them in the problem solving process resulting in better learning.

To answer the second question and verify if the generalized attributes of effective PBL problems were considered by both students and tutors when given sample problems, the students' and tutors' individual responses regarding the effectiveness of eight sample problems from the second phase was analyzed. Results from the present study suggested that the same eleven attributes were referred to by the students and tutors when considering specific problems and problem in general. There were no new attributes generated when referring to specific problems. Implication of this result is that the eleven attributes may be used to assess the effectiveness of specific problems as well as problems in general. Hence we used the eleven attributes as criteria to frequency-score students' and tutors' responses with regard to the effectiveness of the sample problems.

To answer the third question of whether students and tutors differed in their perceptions of the overall effectiveness of the sample problems, the frequency-scores recorded for students and tutors based on their responses about the effectiveness of the sample problems were compared. Despite the different roles played by the students and tutors in the students' learning process and the difference in their expertise, it is surprising that there was no significant difference between the students and tutors in their judgments regarding the overall effectiveness of the problems. A possible reason could be that both groups were engaged in the problem solving process. Given that there is frequent communication in the form of feedback from tutors to students and discussion between students and tutor about the students' learning in PBL (Schmidt 1983), the two groups could have noted similar elements of the problems that influence the students' learning. Kingsbury and Lymn (2008) showed that both the students and tutors agreed on the quality of PBL problems used in a module when evaluating a new curriculum. However, they had explored the problem quality at the program level and not at the individual problem level as in this case. The consensus between students and tutors suggests that feedback from both students and tutors about problem effectiveness could be useful to improving problems.

Amongst the eight problems, P4 was considered to be the least effective problem by both the students and tutors. The most striking feature of this problem is its length at two and half pages– in contrast; the other problems are less than a page in length. Not surprisingly, both students and tutors cited length as a negative attribute of the problem. Both the groups also mentioned that lack of clarity in the problem text made it challenging to identify the intended learning issues. Yet, P4 was considered to be effective in the sense that it made students think and reason, thereby stimulating critical thinking. As for the most effective problem, both the students and tutors cited P2. Problem format, applicability and relevance (to other modules) of the problem, and problem leading to formulation of learning goals were the reasons cited for the effectiveness of P2, whilst difficulty level was considered as a reason for the ineffectiveness of P2. The results suggest that the each of the eleven identified problem attributes may determine problem effectiveness to a varying extent. As a next step, it will be interesting to examine if modifications of the problems based on the participants' judgments leads to an improvement in the effectiveness of the problems. For instance, the participants' responses provided clues that modification of P4 by summarizing and simplifying the problem text will make it more effective.

Finally, to answer the fourth question of whether the eleven attributes in fact related to the effectiveness of the problems, the participants' judgments of the sample problems represented by the frequency-scores were correlated with the student grades. There is, however, one limitation in correlating judgments of problem effectiveness and grades. A correlation measure between the perceived problem effectiveness and the grade can not only be interpreted as the problem judgments reflecting the grade, but it can also be interpreted as the problem judgments being grade-driven. That is, a problem is rated better as a result of getting a higher grade. As students are directly impacted by the grades whilst tutors are relatively unaffected by the grades, tutors' judgments were considered less likely to be biased. Hence we preferred to use the tutor judgments to correlate with the grade. The high and significant correlation between the tutor judgments and student grades suggest that the eleven attributes are indeed associated with the effectiveness of the problems. However, one problem with the present study is that the number of tutors in this study was only five. Hence, we extended the second phase of the study to a different set of eight problems from another first-year module called "Cognitive Process and Problem Solving I". The extended study involved a different group of participants consisting of 18 tutors and 15 students. All other protocols and analysis procedure remained the same. The results (not shown here) suggested not only the repeatability of the study and confirmation of a high and significant correlation between the tutor judgment and student grades, but it also showed the generalized use of the eleven attributes in relating to the effectiveness of problems from different modules.

In summary, this study explored both the student and tutor perceptions about effective problems in general and when given specific problems, and in this process identified eleven attributes. Assessment of the effectiveness of sample problems using the eleven attributes as criteria suggested that the students and tutors agreed with each other on which problem was effective. This consensus correlated well with the students' grades, supporting the conclusion that the eleven attributes are related to the effectiveness of the problems. There are, however, a few limitations to this study. One limitation is that this study does not shed light as to what extent each of the problem attributes influences the students' learning. A second limitation is that this study is retrospective. Utility of the identified problem attributes as criteria to predict quality of untried problems remains to be investigated.

When compared with other studies in literature, this study seems to be the first to collate a list of attributes associated with effective problems based on both students' and tutors' perceptions. Other studies use only the students' perceptions (Dolmans et al. 1995; Jacobs et al. 2003; Soppe et al. 2005) or the tutors' perceptions (Des Marchais 1999), but not both. In addition, this study seems to be the first to consider specific problems and problems in general. Other studies have focused on either specific problems (Dolmans et al. 1995; Jacobs et al. 2003; Soppe et al. 2005) or problems in general (Des Marchais 1999; Kingsbury and Lymn 2008); but not both. This study also attempts to extend beyond identifying the attributes by relating the eleven identified attributes of problems with the students' grades. Despite the association of the quality of problems with the students' academic achievement, with the exception of Soppe et al. (2005), most studies that focus on the attributes of PBL problems do not relate the attributes to academic achievements. Overall, this study has identified eleven problem attributes, derived from both student and tutor opinions, which seem be related to student learning. A feasible follow-up study is to develop a problem quality questionnaire based on the eleven problem attributes and validate it to gain a deeper understanding of the role of the eleven problem attributes in problem effectiveness. This will provide a better insight as to what extent each of the identified attributes leads to overall problem effectiveness and how these attributes are interrelated.

Acknowledgments This research was supported by Republic Polytechnic, Singapore. We gratefully acknowledge the polytechnic for the support rendered and the participants for their involvement in this study.

Appendix

Problem 02: life in a cell

Biological cells operate as independent units capable of managing their internal processes as well as imports, exports and ensuring survival and continuity, just like the way various self-managed entities operate.

Despite the fundamental similarity of all biological cells, different types of cells are able to perform different special functions. Liver cells store glycogen and heart muscle cells exert large forces, whereas red blood cells have no nucleus.

Examine the configuration of biological cells in relation to their ability to self-manage their systems and functionalities.

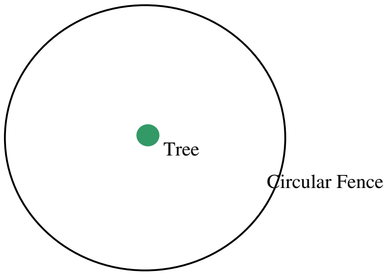
Problem 04: curious spots

A long time ago, and in a land far away, there was a village where a people known as the Curions lived. These people loved to reason things out and explain nature around them, but also had a strong religious belief in a certain god and practiced a curious religious ceremony. Detailed records of their religious experience have now been found, and it is from these records that the following account is pieced together. There is, however, a mystery left behind by the author of the records who was named Augustine.

Every month, during the first night of the new moon when it was pitch dark, the whole village of Curions would make a long journey to a certain deserted place where there lay a magical tree. They believed that the god-spirit of the tree would show its favour on them by making tiny spots appear on the fence surrounding the tree. These spots were so small that they could only be seen in complete darkness, which was why they made their religious observance only at night and at the time of the new moon.

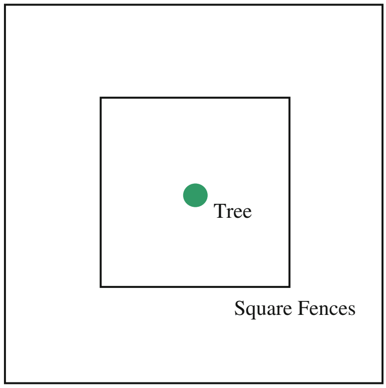
Over many generations, the Curions had tried in vain to look for a pattern in the way the spots appeared. They have since concluded that the spots appeared at random places along the fence determined by the will of the wise god.

However, the elders among the Curions would recount past times, when the fence around the tree was shaped in a perfect circle centred around the tree (see diagram below), how they had carefully counted and totalled up the number of spots appearing for the entire ceremony on each of the wooden sticks that made up the fence. Their fence was built from sticks with exactly the same width.



Invariably, with this circular fence, there would be an almost equal number of spots that appeared on each of the sticks. They believed that this randomness combined with unusual evenness in the appearance of the spots showed that the tree god was not only wise but also was showing its favour on them.

More recently, the Curions had replaced the circular fence with two fences to better protect the tree. These fences were both square in shape, one within the other, and both had the tree exactly in the centre (see diagram below).



With this arrangement, they found that by the end of the ceremony each of the four sides of the inner fence would receive the same number of spots. However, there would no longer be the same number of spots seen for the sticks on one side. There would be a greater number in the middle of each side, and a smaller number towards each of the corners of the square.

On one occasion, the Curions discovered that one of the sticks on the East side of the inner fence had fallen off, so as to leave a gap. While waiting to see what would happen, they discussed among themselves what they might expect. Some of the younger Curions, who were less religious, reasoned that only the sticks that were directly behind the gap (from the perspective of the tree) would have spots appear. However, the elders felt that the tree god had great wisdom beyond their understanding and were not so certain of this outcome.

They soon discovered that the elder ones were proven right. Spots began to appear at all sticks along the East side of the outer fence. Among the sticks on the East side of the outer fence, those that were closest to the gap had the most spots appearing, with a gradually decreasing number of spots that appeared both northwards and southwards of those sticks. Not only that, but to their amazement, some spots also appeared on the North and South sides of the outer fence as well! The subsequent months proved that this was not a fluke, as the same result occurred each time.

A few months afterwards, during a new moon which happened to coincide with a New Year's Day, a second stick was noticed to have fallen from the East side of the inner fence. As it turned out, the two gaps in the fence happened to be symmetrically situated about the centre of the East side. The Curions eagerly watched to see how the spots would appear this time.

The outcome did not disappoint the elders, who by now had concluded that the mystery of the tree god's ways was beyond comprehension. The result of the count showed that there was a clear pattern, with a number of sections showing a high number of spots, alternating with sections showing fewer or no spots. More interestingly, comparing what happened before and after the New Year, for some of the sticks the number of spots appearing was so low that it was even less than the number of spots that appeared before the New Year, when there was only one gap!

As a result of these experiences, almost the entire village had strengthened their belief in this religion, and had come to the conclusion that the manner of appearance of the spots could only be explained by the existence of the wise spirit of the tree. Augustine, however, stated that he had discovered a rational explanation for all that happened, and that this was explained in great detail in another document. Sadly, that document now appears to have been lost to us today.

Nevertheless, according to other Curions who had read that document, Augustine's theory was that the tree first gave out what he called a 'chance wave'. Planks which received more of the 'chance wave' had a higher chance of receiving the next spot given out by the tree. So even though each spot would arrive at the various planks in a lump (just like a particle materializing out of nowhere), the way in which the spots were distributed over a long time would look very much like the way a wave would travel to the various planks.

Your challenge is to examine the account of the experience of the Curions and to decide whether Augustine could indeed have had a logical explanation for all their observances apart from the belief in a tree god.

References

Alwis, W. A. M., & O'Grady, G. (2002). *One day-one problem: PBL at the Republic Polytechnic*. Paper presented at the 4th Asia-Pacific conference on PBL, December 9–13, Hatyai, Thailand.

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem based learning: An approach to medical education*. New York: Springer.

Chai, J. C., & Schmidt, H. G. (2007). *Generalizability and unicity of global ratings by teachers*. Paper presented at the International Problem-Based Learning Symposium 2007, Singapore.

Das, M., Mpofu, D. J. S., Hasan, M. Y., & Stewart, T. S. (2002). Student perceptions of tutor skills in problem-based learning tutorials. *Medical Education, 36*(3), 272–278.

Des Marchais, J. E. (1999). A Delphi technique to identify and evaluate criteria for construction of PBL problems. *Medical Education, 33*(7), 504–508.

Dolmans, D. H. J. M., Schmidt, H. G., & Gijselaers, W. H. (1995). The relationship between student-generated learning issues and self-study in problem-based learning. *Instructional Science, 22*(4), 251–267.

Dolmans, D. H. J. M., Snellen-Balendong, H., Wolfhagen, I. H. A. P., & van der Vleuten, C. P. M. (1997). Seven principles of effective case design for a problem-based curriculum. *Medical Teacher, 19*(3), 185–189.

Gijselaers, W. H., & Schmidt, H. G. (1990). Development and evaluation of a causal model of problem-based learning. In Z. H. Nooman, H. G. Schmidt, & E. S. Ezzat (Eds.), *Innovation in medical education: An evaluation of its present status*. New York: Springer Publishing Co.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235–266.

Hoffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*(6), 722–738.

Hung, W. (2006). The 3C3R model: A conceptual framework for designing PBL problems. *The Interdisciplinary Journal of Problem-Based Learning, 1*(1), 55–77.

Huning, M. (2007). TextSTAT (Version 2.7) [Software]: Huning M.

Jacobs, A. E. J. P., Dolmans, D. H. J. M., Wolfhagen, I. H. A. P., & Scherpbier, A. J. J. A. (2003). Validation of a short questionnaire to assess the degree of complexity and structuredness of PBL problems. *Medical Education, 37*(11), 1001–1007.

Kim, S., Phillips, W. R., Pinsky, L., Brock, D., Phillips, K., & Kaery, J. (2006). A conceptual framework for developing teaching cases: A review and synthesis of literature across disciplines. *Medical Education, 40*, 867–876.

Kingsbury, M., & Lymn, J. (2008). Problem-based learning and larger student groups: Mutually exclusive or compatible concepts–a pilot study. *BMC Medical Education, 8*, 35–45.

Maudsley, G. (1999). Roles and responsibilities of the problem-based learning tutor in the undergraduate medical curriculum. *British Medical Journal, 318*(7184), 657–661.

Morena, R., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory*. New York: Cambridge University Press.

Mpofu, D. J. S., Das, M., Murdoch, J. C., & Lanphear, J. H. (1997). Effectiveness of problems used in problem-based learning. *Medical Education, 31*(5), 330–334.

Schmidt, H. G. (1983). Problem-based learning: Rationale and description. *Medical Education, 17*, 11–16.

Schmidt, H. G. (1993). Foundations of problem-based learning- some explanatory notes. *Medical Education, 27*(5), 422–432.

Schmidt, H. G., & Gijselaers, W. H. (1990). *Causal modeling of problem-based learning*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA, April 16–22.

Shaw, M. E. (1976). *Group dynamics*. New York: McGraw Hill.

Soppe, M., Schmidt, H. G., & Bruysten, R. (2005). Influence of problem familiarity on learning in a problem-based course. *Instructional Science, 33*(3), 271–281.

van Berkel, H. J. M., & Schmidt, H. G. (2000). Motivation to commit oneself as a determinant of achievement in problem-based learning. *Higher Education, 40*, 231–242.