# Characteristics of Problems in Problem-based Learning

Nachamma Sockalingam

# CHARACTERISTICS OF PROBLEMS

# IN

# PROBLEM-BASED LEARNING

*NACHAMMA SOCKALINGAM*

Cover design and layout by N. Sockalingam
Printed by Ultra Supplies, Singapore

# Characteristics of Problems in Problem-based Learning

# Eigenschappen van problemen voor probleemgestuurd leren

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.G. Schmidt

and in accordance with the decision of the Doctorate Board

The public defence shall be held on

Wednesday 20th of January 2010 at 13.30 hrs

by

Nachamma Sockalingam

born in Madurai, India

ERASMUS UNIVERSITEIT ROTTERDAM

## DOCTORAL COMMITTEE

**Promotors:** **Prof.dr. H.G. Schmidt**

**Other members:** **Prof.dr. R.M.J.P. Rikers**

**Prof.dr. H.T. van der Molen**

**Prof.dr. S.E. Severiens**

For Vignes, Krishna,

And

daddy

## CONTENTS

**FOREWORD**

Completion of this dissertation marks the destination of an adventurous journey that I ventured into. As with any journey, this one has been filled with excitement, anxiety, unforeseen challenges, and sense of accomplishment. I have enjoyed the experience and have learnt a few things. I am happy to have reached this destination and to start with a new beginning.

At this juncture, I would like to thank my mentor, Professor Henk Schmidt. This journey would not have been possible without your guidance and support. Dear Prof. Henk, I am grateful to you for the faith that you have in me, for your kind words, and encouragement. Your dedication to go through every word of the papers amazes me and inspires me to try harder. Thanks for everything.

I would also like to thank Republic Polytechnic, with special mention of Mr. Glen O' Grady, Dr. Alwis, Dr. Terence Chong, and Mr. Yeo Li Pheow for having the pioneering vision to initiate educational research in the Polytechnic, and for giving me this opportunity to pursue research work. My thanks to colleagues who joined me in this journey as well: Elaine Yew, Magdeline Lew, Jerome Rotgans, and Judith Williams.

Many others have helped me in my journey – participants, collaborators in my research, colleagues at RP (CED and SAS), and friends. You have inspired me and motivated me. You have cheered me on and supported me. I am especially grateful to Dr. Annet te Lindert, Jovin Hurry, Kiu Cheong Mau, Dr. Noor Shabana, Padma Rao, Paul Chin, and Shawn. Thank you for all that you have done for me. I sincerely appreciate all your support. Thank you.

Last but not least, my thanks to my family: my husband, Sockalingam, and my boys, Vignes and Krishna for their patience, love, and support. Thanks also to my mother and brother for the moral support. Special thanks to Parames for sacrificing time and effort to take care of my boys so that I can focus on my work and research.

I also hope that this work has a ripple effect of inspiring others; especially my students – to aim and achieve what they want to.

*The journey continues…*

# CHAPTER 1

## INTRODUCTION

This dissertation attempts to provide insights about characteristics of problems in Problem-based Learning (PBL). There are five reasons for this dissertation. The first reason is that although PBL has been around for nearly 50 years now, there are still questions about its effectiveness. One of the proposed strategies to examine the effectiveness of PBL has been to understand why and how the underlying principles of PBL (learning is contextual, constructive, collaborative, self-directed and engaging) work, and in under what circumstances (Dolmans, De Grave, Wolfhagen, & Van der Vleuten, 2005; Mamede, Schmidt, & Norman, 2006). To understand how and why PBL works, it is essential to know more about the foundational elements of PBL such as problems, tutors, and students. The second reason is that several studies suggest the quality of problems to be important in positively influencing the students' learning (Gijselaers & Schmidt, 1990; Van Berkel & Schmidt; 2000). In fact, it is shown to have a higher influence on students' learning than tutors' role and students' prior knowledge (Gijselaers & Schmidt, 1990; Van Berkel & Schmidt; 2000). The implication of these findings is that understanding how to improve the quality of problem is likely to be meaningful and crucial to enhancing the students' learning.

The third reason is that despite the suggested importance of quality of problems in PBL, there are only few studies available that shed light on this matter (e.g., Dolmans, Gijselaers, Schmidt, & Van der Meer, 1993; Mpofu, Das, Murdoch, & Lanphear, 1997). What this means is that though we recognize the importance of problems, we do not know much about the characteristics of problems that determine the quality of problems. The fourth reason is that problem designers find the process of designing problems to be challenging

(Angeli, 2002), and the lack of framework based on empirical evidences to guide them indicates that there is not much support for problem designers. Hence, there is a need to gain a better understanding about problems in PBL to help the problem designers, so that they can help students learn better. The fifth reason is that there is a lack of validated instrument available to assess the quality of individual problems in PBL (e.g., Soppe, Schmidt, & Bruysten, 2005; Marin-Campos, Mendoza-Morals, & Navarro-Hernandez, 2004). Hence, assessing the quality of problem may not be straightforward. Therefore, there is a need to develop and validate an instrument to assess the quality of problems so that we can evaluate and design better problems. Underpinning all the above five reasons is the vision to help students learn better in PBL.

The objective of this dissertation, therefore, is to explore the students' and tutors' perspective of problem characteristics, develop and validate a rating scale to assess the quality of problems, and utilize this rating scale to investigate the influence of problem characteristics on students' learning. More specifically, it investigates (1) The students' perspective of characteristics of problems in PBL, (2) Whether students and tutors share a common understanding about problem characteristics, (3) Whether students and tutors actually consider the problem characteristics when evaluating specific problems, (4) Whether students' and tutors' ratings of the problems correspond with the students' grades, (5) Whether it is possible to develop a rating scale to assess these problem characteristics, (6) Whether it is possible to validate and test the reliability of such a rating scale, and (7) Whether such a rating scale be used to assess the influence of problem characteristics on students' learning. To this end, five studies were carried out using both explorative and quantitative approaches. These studies (Studies 1 to 5) are presented as Chapters 2 to 6 of this dissertation.

To present the rationale behind the various studies in this dissertation, this Chapter, Chapter 1, introduces what we know from literature about (1) Constructivist learning, (2) PBL, and (3) Problems in PBL. The literature review gives an overview of not just problems in PBL but also related topics like PBL and constructivist learning so as to provide further insights. For instance, understanding the literature on effectiveness of PBL sheds light on what needs to be explored in improving the students' learning. Although experts reading the dissertation may find the theoretical background as supplementary, I felt that providing the background information will be useful for the other group of my dissertation readers who maybe fairly new to PBL. In doing so, the literature review helps to surface the gaps in understanding about problems in PBL which serve as the guidelines for the design of the five studies presented in Chapters 2 to 6. The implications and limitations of the five studies are discussed in the last Chapter of this dissertation, Chapter 7.

## LEARNING AND CONSTRUCTIVISM

Consider the following newspaper clippings.

There is high chance that flipping through the pages of newspapers on any random day would reveal reports of several "problems" around the world. These problems may be issues about sinking economy, plunging stock markets, soaring unemployment, or declining birth-rates, requiring resolutions. Reading these reports, one may face the problem of anxiety. Yet, the earlier mentioned list is only a sample and not exhaustive. Overall, problems seem to be prevalent in almost every facet of our life. Given the omnipresence of problems in our lives, where and how do we learn to understand and solve these problems? Do we learn to solve problems in school?

Traditionally, a common notion has been that we learn to solve problems in the real-world and not in school. School has been considered to be the place for teaching and learning content knowledge. Mark Twain once said, "I have never let my schooling interfere with education" implying that his education took place in the real-world, outside school. Since Mark Twain's days (1910), times have changed and our views about learning have evolved. As a result, significant educational reforms have taken place. Attempts are now made to bring the real-world to the classroom. So, how do the new approaches help prepare students for the real-world? To understand this better, it is worth taking a look at the educational reforms and the rationale.

According to Mayer (1992, 1999), three views of learning have emerged in the last 100 years. These are learning as response to stimulus, learning as knowledge acquisition, and learning as construction of knowledge. He explains that the view of learning as a response to stimulus became prevalent in the first half of the 20$^{th}$ century, largely based on experimental studies about animal learning. In this view, if the learner responds to a stimulus in a certain manner, then the learner is considered to have learnt. The instructor's role in this case is to provide the stimuli for the learner to respond to, and follow up with immediate feedback. The focus of learning is to learn to respond to stimulus.

Eventually, further studies on animal and human learning in experimental settings gave way to the idea that learning is about "knowing" information and not just responding to stimulus. This view became more common in the mid 20$^{th}$ century. In this second view, learning is seen as acquisition of information and storing of the information in the long term memory. As such, knowledge is seen as a commodity which can be transferred from the instructor to the learner. Therefore, students are first "delivered" information expected of them to know by means of direct instruction, lectures, and well-structured situations. Following that, they are given questions or quizzes to test if they know the information (Uden & Beaumont, 2006; Woods, 1994).

With more studies on human learning in realistic settings, this second view has been redefined and learning is now considered as extending beyond knowledge acquisition to include knowledge construction. This view, also known as constructivism has been gaining wider acceptance since the 1980s and 1990s. According to constructivism, the learner is actively involved in the sense making process, assimilating, and integrating new information, based on what is already known (prior knowledge), to synthesize a mental model or new understanding of what is being learnt. In this case, the instructor provides the students with authentic tasks which represent situations in the real world and guides them when needed. Students work in collaborative teams and interact with fellow teammates as well as the tutor to construct their new knowledge. In addition, they also engage in self-directed learning activities.

In sum, the five underlying principles of constructivist learning are that (1) learning involves knowledge construction, (2) by means of collaborative learning, (3) using authentic, complex, ill-structured tasks that (4) motivate students to (5) engage in self-directed learning (Savery & Duffy, 1995). The focus of learning here is not just the content; but it is to develop a deeper understanding of the topic to be able to transfer the knowledge to a different and varied situation (Mayer, 1999). The constructivist approach is based on the premise that learners must be able to apply and use what they have learnt than just remembering facts.

Wider acceptance of the constructivism is probably augmented by the growing demands of the society. With the globalization and increasing footprint of information technology, employers are no longer just satisfied with employees who know the content knowledge (National Research Council, 1996, The Boyer Commission of Educating Undergraduates in Research University, 1998). They seek employees who are able to apply that knowledge, who are problem solvers, critical thinkers, team players, effective communicators, and resourceful lifelong learners. This is different from the industrial age where workers were mostly needed to carry out instructions/tasks (Reigeluth, 1999). Of the three views of learning, constructivist learning seems to be the one that best meets these changing needs of employers and help foster better learning.

**CONSTRUCTIVISM AND PROBLEM-BASED LEARNING**

To implement the ideas of constructivist learning, educators are required to redefine teaching and learning methodologies (Knowlton, 2003). That is, the instructional approach and learning environment need to be aligned with the underlying views of learning (Reigeluth, 1999). This is important because instructional approach and learning environment that may be suitable for knowledge acquisition may not be suitable for knowledge (acquisition and) construction. For example, direct instruction by a tutor to “show” and “tell” how to carry out a skill may result in the student knowing how to repeat these steps. However, the students may not really understand what happens if the steps are not followed in that particular sequence and why or how to modify the steps in different situations. The learning environment is no exception. If students are seated individually like in lecture theatres, collaborative learning is likely to be restricted. Hence, the instructional approach and learning environment needs to support the underlying principles.

One instructional approach grounded and adapted to the principles of constructivist learning is Problem-based Learning (PBL). As the name suggests, PBL revolves around problems. Problems are carefully designed descriptions of problematic situations which require explanations or resolution. The problems typically represent real-world situations and hence are purported to prepare the students in understanding and solving such problems. Students work in collaborative teams to resolve these messy or ill-structured problems (Jonassen, 1997; Savery & Duffy, 1995; Voss & Post, 1988). To do so, they brainstorm and discuss the problem to first identify what they know, do not know, and need to know about the problem. Elaboration during the group discussion is suggested to activate students' prior knowledge and trigger their interest.

This elaboration helps to surface different views and even misconceptions held, eventually helping students to identify what they do not know. The gaps in their understanding are thought to stimulate epistemic curiosity (Berlyne, 1978). The learning issues identified serve to guide the students in their self-study. When in doubt, they may consult with the tutor. Following the period of self-study, the students reconvene to share their findings and discuss these and eventually co-construct their shared understanding of the issues presented in the problem. Overall, this process of problem-solving in PBL is founded on the earlier mentioned principles of constructivist learning. The underlying belief in this student-centric constructivist approach is that it will better prepare students for the real-world.

Despite the promise of PBL, the practicality of the underlying assumptions needs to be questioned. For instance, one assumption is that problems simulating real-world scenarios can be designed. A second assumption is that students are willing to tackle a given problem. A third assumption is that the confrontation with messy or ill-structured problems results in learning. These assumptions, however, are theoretical and require validation. Several questions can be raised. Can we design authentic problems? How should authentic problems be? Are students motivated to tackle these problems? Do ill-structured problems result in ineffective learning? Such questions serve as the motivation factor for this dissertation. The objective of this dissertation is to address some of these questions about problems used in PBL. To this end, the subsequent sections of this Chapter reviews literature on PBL and problems in PBL. Finally, gaps in our understanding about problems in PBL are identified, and studies to address these gaps are proposed.

## WHAT IS PROBLEM-BASED LEARNING?

Several definitions of PBL exist in the literature. According to one of the earliest definitions, PBL is the process of working towards the understanding and resolution of a problem and involves acquisition of professional skills (Barrows & Tamblyn, 1980). Vernon and Blake (1993) propose that PBL is an instructional approach that uses problem as a context for student to acquire both problem-solving skills and knowledge. Uden and Beaumont (2006) add that PBL includes problem-solving activities, critical thinking exercises, collaborative learning, and independent study, allowing students to relate to the context of the problem and construct new meaning. Schmidt (1983) tries to clarify the differences and proposes that the variation in definitions is the result of the various interpretations of PBL.

In a latest report, Schmidt, Van der Molen, and Te Winkel (2009) suggest that the various descriptions of PBL can be grouped into three types, depending on the learning objectives of the PBL. The first type of PBL, type I PBL, is defined as information-processing or cognitive constructivism approach (Hmelo-Silver, 2004; Norman & Schmidt, 1992; Schmidt, 1983; Schmidt, 1993; Schmidt, De Grave, De Volder, Moust, & Patel, 1989; Schwartz & Bransford, 1998).

The idea behind type I PBL is that it helps students *to make sense of the world by constructing dynamic mental models* about the information embedded/presented in a problem through group discussion and self-study. In the process of problem analysis, students formulate a tentative hypothesis about the situation described in the problems based on what they know. The assumption here is that students have sufficient prior knowledge to analyze the problems, even if some of these are misconceptions. The gaps in their understanding and identification of these gaps is believed to trigger their epistemic curiosity (Berlyne, 1978), which engages them in collaborative and self-directed learning. Discussion in the groups with the fellow students and tutors is thought to result in formulation and reformulation of new understanding, in addition to motivating them. This view of PBL is founded on the cognitive psychology principles that learning involves (1) activation of students' prior knowledge, (2) elaboration of content, (3) restructuring of semantic networks/schemata, and (4) development of an intellectual scaffold (Schmidt, 1993).

The type II PBL, characterizes it as a *problem-solving process*. In other words, PBL is defined as a process of inquiry. The goal of PBL in this case is to learn diagnostic reasoning and to mimic the thought process of the expert (rather than knowledge acquisition). For instance, Barrows (1986) suggests that the primary educational objectives of PBL is to help medical students develop effective clinical reasoning skills, organize knowledge base for use in clinical context, develop self-directed learning skills, and be motivated to learn. This PBL definition is more common in some of the medical curriculum adopting PBL (Barrows & Tamblyn, 1980; Barrows, 1990). However, research shows that general reasoning skills may not be that transferrable (Barrows, Neufeld, & Feightner, 1978; Elstein, Shulman, & Sprafka, 1978; Neufeld, Norman, Feightner, & Barrows, 1981). Furthermore, with the adoption of PBL in non-medical schools, reasoning skills and knowledge construction can be viewed broadly. Hmelo-Silver (2004) refers to this as constructing an extensive and flexible knowledge by integrating information across multiple domains.

According to the cognitive psychological model, professional competence development is a transition from conceptually rich and rational knowledge base to a non-analytical ability to handle situations efficiently and effectively. This is based on the findings that experts tend to have minimal explicit reasoning; instead they rely on implicit reasoning in which the reasoning process is automated and the cognitive processes are condensed into pre-programmed “scripts” (Van der Vleuten, 1996). Schmidt, Norman, and Boshuizhen, (1990) refer to these scripts as “illness scripts” in the medical context (Also see Custers, Boshuizen, & Schmidt, 1998). This view is supported by findings from research that expertise in clinical reasoning is associated with knowledge encapsulation and flexible use of knowledge (e.g., Boshuizen & Schmidt, 1992). Hence even though the objective of type II PBL is to teach problem-solving skills, what seems to be needed is a strong knowledge base emphasized by type I PBL.

Finally the third type of PBL, type III PBL characterizes it as an instructional method fostering skills which teach students *how to learn* (Silen & Uhlin, 2008; Toon, 1997). This type of PBL is founded on the belief that knowledge is tentative and ever changing, and therefore it is important to know how to learn than what is learnt. Although PBL students have been found to make more use of the library resources, there is no evidence that these skills foster life-long learning skills. Another perspective is that these learning skills can be regarded as tentative as well. For instance, looking back at history, before the invent of technology tools such as internet, (Google scholar and PubMed), the skills needed to access information from hard copy journals were different from that required for using today's technology tools. In the same line of thought, leaning skills needed in future may be different from what students are accustomed to at present. If the idea in type III PBL is that knowledge is of lesser importance as it is tentative and changing with time, then it is reasonable to question the importance of skills associated with learning how to learn as they are tentative too.

Despite the different perspectives on PBL, there are some common elements in the implementation of PBL. These are (1) the use of problems to initiate the learning process, (2) collaborative work, (3) flexible guidance from a tutor, (4) minimal lecture, (5) self-directed learning by students, (6) ample time made available for collaborative work and self-study (Schmidt et al., 2009). More details on some of these elements of PBL such as problems, students, and tutors are provided in a later section on "Foundational elements of PBL".

## HISTORICAL BACKGROUND OF PBL

Philosophy of PBL is thought to date back to Dewey (1916) who believed that engaging and experiencing results in learning. In his words, "Methods which are permanently successful in formal education go back to the type of situation which causes reflection out of the school in ordinary life. They give pupils something to do, … and doing is of such a nature as to demand thinking, or the intentional noting of connections; learning naturally results" (Dewey 1916, 1944, p 154). This resonates with the modern principles of constructivist learning that learning should be contextual, constructive, self-directed, collaborative, and engaging (Dolmans, et al., 2005; Mayer, 1999).

The first implementation of PBL is reported to be in McMaster University (Canada) in the 1960s (Hamilton, 1976; Neufeld & Barrows, 1974). Traditionally, learning in medical schools has been mediated through information-loaded lectures, delivered by content experts to a mass of students. This approach to learning is rather disconnected from the real-world practice of medicine. Although students who learn through the traditional curriculum were found to be well-prepared in terms of factual knowledge, they were found to be limited in the application of the knowledge (e.g., study by Gonella, Goren, Williamson, & Cotasonas, 1970). To address this, the PBL approach was adopted. Students were to learn primarily through self-directed study guided by the problems. In addition, the curriculum aimed to instil professional skills such that they are able to integrate knowledge from different domains, make decisions, work well with others, and effectively communicate with patients (White 1996).

Following McMaster, several other medical schools including University of Limburg at Maastricht (Netherlands), the University of New Mexico (United States), and University of Newcastle (Australia) implemented variations of PBL (Uden & Beaumont, 2006). Currently, 80 percent of the medical schools in the United States (Jonas, Etzel, & Barzansky, 1989) and over 60 medical schools across the world (Delisle, 1997) practise PBL in whole or to some extent. In addition, PBL has been implemented in various other fields of study such as architecture, law, engineering, and social work (Boud & Felleti, 1991). PBL has also been adopted in various academic levels ranging from kindergarten to higher secondary schools in United States (Torp & Sage, 2002). The slow but steady growth of PBL since the 1960s suggests that despite the scepticism about PBL in the initial stages, it is now gaining popularity. This popularity is also seen in the modified application of PBL which is explained below.

## OTHER CONSTRUCTIVIST APPROACH RELATED TO PBL

Modifications of PBL are seen in other forms of constructivist learning such as project-based learning, case-based learning and inquiry-based learning. Project-based learning is similar to PBL except that it is driven by a complex project consisting of many cases which extend over a longer period of time (Hmelo-Silver, 2004; Krackjick, Blumenfeld, Marx, & Soloway, 1994, Savery, 2006). Case-based learning is driven by a case which is usually a narration of an authentic case of legal, medical or social-work nature. However the case is often presented only after other instructions (Savery, 2006; Williams, 1992). In inquiry-based learning (IBL), learning process is triggered by a question. In addition, the key difference between IBL and PBL is that in the former, the tutor is both a facilitator and provider of information; In PBL, the tutor does not provide the information (Savery, 2006). Despite the differences in various approaches to constructivist learning, one key common element amongst the different approaches is the use of authentic, ill-structured instructional materials (Jonassen, 1999). This is referred to as “Problems” in PBL.

## EFFECTIVENESS OF PBL

The widespread adoption of PBL could be attributed to the effectiveness of PBL. However, Sanson-Fisher and Lynagh (2005) question whether the prevalence of PBL is due to successful dissemination, rather than the demonstration of positive educational outcomes. Several studies have tried to answer this question by comparing PBL curricula with traditional curricula based on curricular level outcomes such as knowledge acquisition and knowledge application. For instance, Dochy, Segers, Bossche Van den, and Gijbels (2003) showed that PBL students are better in knowledge application. However, they did not find significant difference in students' knowledge acquisition. Norman and Schmidt's study (2002) reviewed the available evidences to evaluate the theoretical advantage claimed for PBL and concluded that there was a higher transfer and retention of knowledge, enhancement of self-directed skills, and a higher level of learner satisfaction in PBL. Albanese and Mitchell (1993) showed that PBL students reported higher level of satisfaction (also Colliver, 2000; Vernon & Blake, 1993), and better preparedness for independent learning. Overall these studies seem to support the use of PBL curricula.

Nonetheless, researchers such as Colliver (2000) and Newman (2003) showed in their meta-analysis that there is no evidence for the superiority of the PBL curricula over traditional curricula on the basis of conventional knowledge tests and question the effectiveness of implementing PBL. In line with this, others put forward the idea that PBL is minimally guided, and that minimally guided instructions are incompatible with the cognitive structure of human mind. Therefore, they suggested that the outcomes of PBL curricula are likely to be sub-optimal (Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004).

Analyzing these earlier mentioned studies, Mamede, et al. (2006), and Norman and Schmidt, (2000) argued that the use of curriculum level randomized controlled studies advocated by Colliver (2000) and Newman (2003) are not suitable method to measure the impact of educational interventions. Although these authors acknowledged that use of experimental studies, they questioned the feasibility and applicability of large-scale randomized curriculum level studies with reference to the question asked. They reasoned that the large-scale randomized curricular studies are not feasible for three main reasons.

The first reason is that there are number of other variables/components that are involved in the implementation of PBL and traditional curricula other than just the difference in the curricula. As it is not possible to control these various variables, the complex interactions amongst them could result in confounding results, hence making it not possible to establish the interrelationship between the variables or to conclude that the learning outcomes are due to the curricular "treatment". The second reason is that even if these confounding factors are known, for example, profile of students, it is difficult to randomize across educational institutions. For instance, students admitted in the various educational institutions may go through different admission selection criteria. The third reason is that blinded treatment of curricula is not possible as students and tutors are both aware of the curricula they are engaged in. For all these reasons, they argued that large-scale randomized curricular level comparative studies are not feasible in the first place and advocate that the research design is guided by the questions asked by the research.

Taking these factors into consideration, Schmidt et al. (2009) carried out a meta-analysis comparing students and graduates from a Dutch medical university that uses PBL curriculum with their counterparts from other Dutch medical universities using traditional curricula. The key difference between this study and the earlier curricular comparative studies are that (1) one well-established PBL curriculum was compared with other traditional curricula, rather than comparing several PBL curricula with other traditional curricula, (2) several educational outcomes such as medical knowledge, clinical reasoning, drop-out rate, graduation period, interpersonal skills, and the satisfaction level in curriculum were compared rather than focusing on the knowledge tests, and (3) participants were all of similar profile in terms of prior performance and previous learning experience.

Their results suggest that the students in the PBL curriculum were better in terms of interpersonal skills. They also expressed higher level of satisfaction which was reflected in the lower drop-out rate and shorter time for graduation. In addition, medical knowledge and diagnostic reasoning were found to be slightly better in the PBL curriculum. These outcomes demonstrate that PBL curriculum has positive effects in students' learning. The decreased drop-out rate and shorter graduation time suggests that PBL is efficient and has positive implications for the implementation of PBL.

An additional inference from these outcomes is that even with minimal direct instructions, students in PBL curriculum demonstrate better learning in terms of knowledge and especially in terms of interpersonal skills. As to the reason why PBL is able to achieve this even without direct instruction, Schmidt et al. (2009) offered an explanation that this function is taken over by the problems used, collaborative learning with fellow students, and formative feedback from tutors which provide instructional scaffolding to support the students' learning in a more holistic manner. One thing to note is that the PBL curriculum compared in this study was the type I curriculum and repeatability of these results in type II and type III curriculum remains to be tested.

Another line of argument is that comparison of PBL with traditional curricula is futile (Dolmans, et al., 2005) as the comparison of PBL with traditional curricula is analogous to comparing apples with oranges. Since the underlying principles of PBL and traditional curricula are different, the comparison of the two learning approaches is not considered appropriate (Dolmans et al., 2005).This has led to calls for more research to understand why and how the underlying principles of PBL (learning is contextual, constructive, collaborative, self-directed and engaging) work, and in under what circumstances (Dolmans et al., 2005; Mamede, et al., 2006).

## FOUNDATIONAL ELEMENTS OF PBL

### *Problems in PBL*

To understand how and why PBL works, it is essential to know more about the foundational elements of PBL. According to Majoor, Schmidt, Snellen-Balendong, Moust, & Stalenhoef-Halling (1990), the three key elements of PBL are problems, students, and tutors. Problems used in PBL contextualise real-world problems and are typically a set of descriptions of phenomena or situations in need of explanations and resolution (Schmidt, 1983). They are often presented in textual format, sometimes with illustrations, pictures, videos, and simulations (Schmidt, 1983; Schmidt & Moust, 2000). Problems are sometimes classified as explanation problems, description problems, or strategy problems based on their objectives (Schmidt & Moust, 2000). An example of an explanation problem is given below.

*Example of an Explanation Problem from a Cognitive Processes and Problem Solving Skills Module*

**Education, what is it?**
Ivan Pavlov was a Russian biologist who received the Nobel Prize in 1904 for Medicine. He found out during a study that every time a bell is sounded when a dog is given food, the dog would salivate. Eventually, the dog would salivate even when just the bell rang without food.
Psychologists who had defined learning as what causes a "change in behaviour" concluded that the dog has learned something which it

could not do before. This happening of “learning” in the dog has since become a famous example of “classical conditioning” in the so-called Learning Theory.

Sceptics criticize that if we link learning to change in behaviour, then if someone suffered a leg injury and started to limp, it would be acceptable to say that the injured person had learned to limp.

Quite clearly, there is so much confusion about learning. However, the more important question to individuals, communities and tax-payers, is about education rather than learning. Some people believe that learning is the same as receiving an education, yet many would be unwilling to consider that the Pavlov’s dog got educated to salivate, or someone got educated to limp following an injury.

What could be meant by the phrase “receiving an education”? What makes someone “educated”?

Problems are the starting point of students’ learning process in PBL. These problems try to actualize real-life situations that students will have to explain in their own words. Students begin working on the problem with no preliminary preparation but just their prior knowledge. They work in their collaborative teams to analyze the problem and try to explain the phenomenon described. In this process, they come up with a tentative hypothesis about the problem.

In answering the earlier mentioned problem, students may recognize from the problem that learning is considered to be more than responding to a stimulus. However, from their personal experience, they may associate learning with going to school to get an educational certificate. Or they may propose that the more information one knows, the more educated the person is. Yet another may counter this notion by hypothesizing that knowing more information may not mean that the person can apply this knowledge. Such discussion raises several questions which may lead to proposition of tentative conclusions. One example of a tentative conclusion could be that "learning may not necessarily mean knowing information".

These questions and propositions then serve as the guidelines for the students' self-study. Not only that, the identification of gaps in their understanding engages them in their self-study. As a result, students refer to various resources such as internet, books, and news articles to find out more on these issues based on what they deem as relevant and important to respond to the problem. After the self-study period, the students reconvene with their team to share their findings, explain their views and synthesize a shared understanding about the issues presented in the problem.

Thus, the problems serve to engage the students, 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 (Hmelo-Silver, 2004). Typically, several problems pertaining to a specific subject area are organized in a module. For example, at Maastricht University, a complete module in the psychology curriculum consists of 10 problems (Verkoeijen, Rikers, Te Winkel, & Van der Hurk, 2006)

### *Students in PBL*

Students are another foundational element of PBL. Usually, they work in teams of 6-10, to resolve the given problem (Schmidt, 1983). They work independently during self-study period of the learning process, and collaboratively with their team members during group discussion to construct their own knowledge. Studies suggest that students' prior knowledge (Schmidt, De Volder, De Grave, Moust, & Patel, 1989; Schmidt & Gijselaers, 1990), and conceptions about learning (Loyens, Rikers, & Schmidt, 2006, 2007) impact their achievements in learning.

***Tutors in PBL***

With regards to tutors in PBL, the tutor's role differs from that of teachers in the traditional curriculum. In the latter case, teachers provide information directly to students by means of lecture or teaching (Woods, 2004). However, the PBL tutors do not provide information directly. Instead, they help students to be more self-directed, motivated, and collaborative critical thinkers. Tutors do this by observing students' learning activities, diagnosing issues faced by the students, and intervening at appropriate instances to provide required scaffolding (Das, Mpofu, Hasan, & Stewart, 2002; Maudsley, 1999; Neville, 1999).

These three foundational elements of PBL interact closely with each other in the PBL process. Although contextual differences, such as number of students in a group, number, duration, and frequency of meetings, exist in the implementation of PBL across various educational institutions (Newman, 2003), the three foundational elements of PBL remain common. The next section presents a brief summary of studies focusing on the various elements of PBL.

## RESEARCH FINDINGS WITH REGARDS TO VARIOUS ELEMENTS OF PBL: IMPORTANCE OF PROBLEMS IN PBL

In general, two approaches are undertaken to gain more in-depth understanding of how and why PBL works. One approach is to focus on specific elements of PBL. For example, various studies have been carried out to understand the role of the tutor (e.g., Kokx, & Boon, 1993; Schmidt &, Moust, 1995), students' conceptions of learning (e.g., Steinert, 2004), collaborative learning (e.g., Visschers-Pleijers, Dolmans, Wolfhagen, & Van der Vleuten. 2005), and problems (e.g., Yueng, Ay-Yueng, Chiu, Mok, & Lai, 2003). Another approach has been to look at the interrelationship of the elements (e.g., Dolmans & Wolfhagen, 2005). Both approaches should be seen as complementary. While the first approach provides more detailed information on specific elements of PBL, the second approach provides an overview and interrelationship of the different elements of PBL. Hence, a combination of both the approaches is likely to be useful. An example of the first approach is the study by Schmidt and Moust (1995) which used a causal modelling technique to examine the characteristics of effective tutoring based on students' perceptions. They found that effective tutoring is influenced by three distinct and interrelated qualities: tutors' content knowledge, tutors' willingness to become involved with the students in an authentic way, and tutors' communication skills.

Taking the second approach, Gijselaers and Schmidt (1990) investigated the interrelationship between the three foundational elements, two process elements and two outcome elements of PBL. The foundational elements, otherwise referred to as input elements, are quality of problem, students' prior knowledge, and tutor performance. The process elements are students' individual self-study activity, and group functioning, and the two outcome elements are students' achievement, and interest. They found that among the three input elements, the quality of problems had the most direct and highest influence on the process elements and through these process elements on the outcome elements (See also Schmidt & Gijselaers, 1990, and Van Berkel & Schmidt, 2000). In addition, they found that the highest interaction within the three foundational elements was between the quality of problem and students' prior knowledge.

Extending on these studies, Van Berkel and Dolmans (2006) investigated the interaction between tutor performance, tutorial group productivity and the effectiveness of a PBL unit. They also found that the quality of problems influenced the group functioning which then impacted the achievement. Collectively, these studies investigating the interrelationship between different elements of PBL suggest that the quality of problem plays a significant role in students' learning compared with the students' prior knowledge and tutor performance. The significant role of problem in PBL raises the following questions: (1) what are the characteristics of good problems?, and (2) what evidences are there about individual characteristics of problems in influencing students' learning? So let us now find out about what is known of problems in PBL.

## WHAT IS KNOWN ABOUT CHARACTERISTICS OF PROBLEMS IN PBL BASED ON THEORY?

Angeli (2002) found designing and selecting problems to be the two key challenges faced by high-school teachers in implementing PBL. Gagne, Wager, Golas, and Keller, (2005) suggest that different types of learning and instruction are meant for different learning outcomes and advocates that careful thought must be given to the design of the instruction. Understanding what characterizes good problems is therefore necessary to designing such problems. This raises the questions on how educators have been designing problems. The answer to this is that problems have been designed mostly based on experiential knowledge and principles or guidelines drawn from the learning and cognition theories. For example, Dolmans, Snellen-Balendong, Wolfhagen, and Van der Vleuten, (1997) derived seven principles of effective case design based on learning and cognition theories such as constructivist learning and available empirical evidences. The seven principles of case design are that the case should simulate real-life, lead to elaboration, encourage integration of knowledge, encourage self-directed learning, fit in with students' prior knowledge, is of interest to students, be of an adequate level of complexity and structuredness, and reflect the faculty's objectives. Table 1 shows the parallel between the seven principles and the five constructivist principles.

**Table 1**

*Principle of Problem Design in Comparison with Principles and Theories in Learning*

| Learning domain (Bloom, 1956) | Assumptions of constructivist learning in PBL (Savery & Duffy, 1995) | Seven principles of problem design (Dolmans et al., 1997) | Hung's 3C3R model of problem design (Hung, 2006) |
|---|---|---|---|
| Cognitive | Learning involves construction of knowledge | Reflect the faculty's intended learning objectives | Content Context Connection |
| | | Encourage integration of knowledge | |
| | Learning should be based on authentic or real-world situations | Stimulate real-life scenario | |
| | | Fit with students' prior knowledge | |
| Affective | Learning should engage students | Interest the students | Context |
| Behavioural | Learning should be collaborative | Lead to elaboration | Researching, Reasoning Reflection |
| | Learning should promote self-directed learning | Encourage self-directed learning | |

More recently, Hung (2006) proposed a conceptual framework represented by the 3C3R model for designing problems in PBL. This framework comprises two components: core components and processing components. The core components refer to content, context, and connection which support the content and contextual learning while the processing components refer to researching, reasoning, and reflecting which support the cognitive processes and problem solving Skills. The core components relate to the idea of contextual and motivated learning while the process components relate to the idea of self-directed learning. Several other authors also suggest guidelines and framework for designing problems (Duch, Groh, & Allen, 2001, Torp & Sage, 2002; Uden & Beaumont 2006), which resonate with Dolman's (1997), and Hung's (2006) proposal. Despite the general consensus among the several authors on the guidelines for designing PBL problems, these assumptions about problem characteristics tend to be theoretical and require validation (Dolmans et al., 1997; Hung, 2006). This raises the question on what empirical evidence is present with regards to the characteristics of problems, and the effectiveness of problems.

**WHAT IS KNOWN ABOUT PROBLEMS FROM EMPIRICAL RESEARCH?**

Generally, studies on problem characteristics focus on specific problem characteristics. An exception to this is Des Marchais' study (1999) which used a Delphi study approach with six experts to identify characteristics of good problems. Nine characteristics of good problems were generated, of which the two most important characteristic were found to be (1) the problem should stimulate thinking, analysis, and reasoning, and that (2) the problem should lead to self-directed learning. This study is notable as it was the first study to provide an overview of nine attributes of problems using a systematic approach. As far as known, this is the only empirical study to identify a comprehensive list of problem characteristics. However the limitations are that the study used only the expert's views. Students' views were not included. One argument is that students are the end-users of the problems and therefore their perceptions of the problem characteristics needs to be investigated. In addition, the study does not attempt to assess whether the identified characteristics contribute to the effectiveness of the problem; it is more of an explorative study to identify the characteristics.

Most other studies tend to focus on specific problem characteristics. These studies can be further classified into two categories. The first category of studies try to understand more about the problem characteristics (Jacobs, Dolmans, Wolfhagen, & Scherpbier, 2003; Marin-Campos et al., 2004). The second category of studies go one step further and investigate the influence of problem characteristics on students' learning (Dolmans, et al., 1993; Mpofu, et al., 1997; Soppe, et al., 2005; Verkoeijen, et al., 2006).

In the first category, Jacobs, Dolmans, Wolfhagen, and Scherpbier (2003) developed and validated a questionnaire to assess the degree of complexity and structuredness of PBL problems. They defined "complexity' as "the number of characteristics or variables that play a role in challenging the students to think and learn as well as the interrelationship between these characteristics, and the stability of this interrelationship over time". "Structuredness" of a problem is "characterized as requiring the application of a limited number of well-structured rules, with solutions that are straightforward and predictable". Therefore, a well-structured problem is thought to have one defined solution compared to an ill-structured problem which may have many possible solutions. They found that although students could clearly differentiate between simple and difficult problems, they had difficulty in discerning ill-structured from complex problems. Hence the authors classified both ill-structured and complex problems as "too difficult" problems.

Marin-Campos et al. (2004) developed an 18-item rating scale with the objective to measure three characteristics of PBL problems. The three problem characteristics assessed were (1) the extent to which the problem leads to learning activities such as self-directed study, and collaborative learning, (2) how the problem is structured/represented to lead the students to the intended content, and (3) the extent to which time and resources were needed to tackle the problem. They administered the rating scale to students to gather feedback on 14 problems and found that measurement using the rating scale was reliable.

Studies in the second category go one step further compared with the first category, and investigate the influence of the problem characteristics on students' learning. One of the objectives of PBL is that students' construct their new knowledge as a result of working in the problem. This can be expected to be represented by the learning issues identified by the students as a result of working on the problem. Therefore, to measure the effectiveness of problems, one approach is to evaluate whether students are able to generate the same learning issues as intended by the curriculum. The degree of congruence between the two is considered to be reflective of problem effectiveness (Dolmans, et al. 1993; Mpofu, et al. 1997).

To this end, Dolmans et al. (1993) compared the students-generated and faculty-intended learning issues across 12 problems. They found that, on average, students were able to identify 64% of the intended learning issues across 12 problems. Similar results were also noted by Mpofu et al. (1987). Though these studies are useful attempt to identify the effective problems, a limitation is that only one characteristics of the problem; the extent to which the problems lead to the intended learning issues are considered. In addition, this method is time and resource intensive. For instance, if each problem had an average of 5 intended learning issues, and there were 120 students, then nearly 7200 comparisons have to be made for 12 problems.

Another approach to investigate the influence of other problem characteristics such as problem familiarity (Soppe et al., 2005) and goal clarity (Verkoeijen et al., 2006) on students' learning, in a more efficient manner has been to use a questionnaire/rating scale. For instance, Soppe et al., (2005) used a 12-item rating scale to assess students' perceptions about the level of problem familiarity and interestingness, and compared that with (1) quality and quantity of learning issues generated by students, (2) students' achievement, and (3) time taken for self-study. Two groups of students who were working on either a "familiar" or "unfamiliar" versions of the same problem were administered the rating scale. The results suggested that the students working on the problem with the familiar context indeed perceived it to be familiar and interesting. However, no significant difference was found between the two groups of students in terms of the learning issues generated, achievement, and the amount of time taken for self-study.

Verkoeijen et al. (2006) investigated the extent to which goal specification in a problem influenced students' learning by administering either a "goal-free" or a goal-specified" problem to two groups of students. The goal-free problem did not specify any goal whereas the goal-specified problem stated the goal needed to be achieved by students. The authors postulated that the students working on goal-free problem would read more articles, and spend longer time in self-study. In addition, they expected that these students would spend more time in reporting phase. To measure the learning outcomes, they administered a short rating scale which assessed the quality of the PBL cycle in terms of (1) the depth and quality of discussion and reporting phases, (2) the extent of elaboration during the discussion phase, and (3) the students' perceived mastery in the subject matter. In addition, measures of time taken by students for the different phases (discussion phase, reporting phase, and self-study) and the number of articles were measured.

The results showed that the students working on the goal-free problem read more articles, spent more time in studying as well as reporting the findings. Furthermore, the students working on the goal-free problem perceived to have a higher level mastery of the subject matter compared with the students working on the goal-specified problem. These two studies were unique in attempting to use an experimental approach and in using a questionnaire to relate the problem characteristics with the students' learning. However, the questionnaires used in these studies are not validated. Students' perceptions were also not corroborated with tutors' observation on the same measures.

The value of the earlier mentions studies is that they add to our understanding of problem characteristics, in particular, what defines problem characteristics, the impact of specific problem characteristics, and the ways to assess the effectiveness of problems. They also provide insights on the diversity of research methods useful to answer specific questions. For instance, the results suggest that phenomenological approach such as focus group study or Delphi study maybe useful to identify the problem characteristics while experimental studies are applicable to investigate the influence of a particular problem characteristic on the students' learning. Nevertheless, there are a several limitations in the existing studies on problem characteristics.

First, not many studies use an empirical approach to define a broad range of problem characteristics. Second, studies that define problem characteristics include only the students' perceptions or tutor's perceptions but not both. Third, the relationship between problem characteristics and students' learning are not generally assessed by studies exploring a wider spectrum of problem characteristics. Fourth, some of these studies use resource and time-intensive methods, such as in the case of comparing the student-generated learning issues with the faculty-intended learning issues. Though such research intensive method may provide detailed information, this may not be applicable when the quality of several problems need to be evaluated. Fifth, instruments used to measure the problems characteristics such as Jacobs et al.'s (2003) questionnaire and Marin-Campos et al.'s questionnaire (2005) focus on a selected few characteristics. In addition they are seldom validated (Marion-Campos et al., 2004; Soppe et al., 2003; Verkoeijen et al., 2006). Sixth, students' perceptions about the influence of problem on their learning are not corroborated with other measures.

**SUMMARY OF LITERATURE REVIEW**

In summary, from the literature review on the various elements of PBL, we can draw that the quality of problem plays a significant role in students' learning compared with the students' prior knowledge and tutor's function. However, designing and evaluating problems is found to be challenging. In addition, there is relatively few research on the quality of problems. Though there are some existing studies addressing various issues on problems in PBL, there are number of shortcomings in the existing studies. These shortcomings are elaborated above. Such gaps in our understanding of problem characteristics bring about the following questions

- What are the students' perceptions of effective problems?
- Do students also consider several problem characteristics?
- Do students and tutors share a common understanding about these characteristics?

- Do students and tutors actually consider these characteristics when evaluating specific problems?
- Do students' and tutors' ratings of the problems correspond with the students' grades?
- Is it possible to develop a rating scale to assess a more comprehensive list of problem characteristics than what is available at present?
- Is it possible to validate and test the reliability of such a rating scale?
- Can such a rating scale be used to assess the influence of problem characteristics on students' learning?

These questions served as the motivation and guidelines for the five studies carried out as part of this dissertation.

## OVERVIEW OF STUDIES

### STUDY 1

Since there is a lack of empirical research on characteristics associated with good problems, especially from the students' perspectives, the first study explored the students' perceptions of good problems based on their experiences with problems in problem-based learning. Despite the existence of principles and guidelines which give list of problem characteristics, most of these tend to be theoretical and not validated. The limited few empirical studies tend to focus on specific characteristics. One rare study which has identified a comprehensive list of problem characteristics using Delphi technique is Des Marchais' study (1999). This study identified nine characteristics associated with effective problems. Though useful, this study had not included the students' perceptions. Students are the end-users of the problems and can be considered as novices in terms of the content knowledge in comparison with the faculty. Cognitive psychology suggests that experts and novices process information differently (Van der Vleuten, 1996). Hence, it is possible that tutors' perceptions of the quality of problem are different from that of the students' perceptions. This motivated us to find out the students' perceptions on the characteristics associated with effective problems. To this end, we asked students from a PBL curriculum to reflect and record their perceptions of what makes a good problem in their e-journals. Students' submissions were then text analyzed to identify the characteristics of good problems in PBL. Results and conclusion of Study 1 are presented in Chapter 2.

**STUDY 2**

Study 2 builds on Study 1 and investigated both the students' and tutors' perceptions of effective problems by means of focus group studies. The objective of this study was to first compare the students' and tutors' perceptions of effective problems and determine if there is any qualitative difference. As far as we know, most of the studies in PBL literature focused on either the students' perceptions or the tutors' perceptions of problems. One of the few exceptional studies is the likes of Dolmans et al.'s (1993) study which compared the student-generated learning issues with the faculty-intended learning issues. However, these studies only focused one problem characteristic; the extent to which the problem led to the intended learning issues. In addition, we also wanted to find out if both these groups held their perceptions when given specific set of problems and whether their ratings of problem correlates with the students' grades. As mentioned previously, it is possible that tutors view the effectiveness of problems differently from the students. Furthermore, asking participants to mention desirable characteristics of problems in general may yield different answers from asking them to mention characteristics of specific problems. Finally, with the exception of Soppe et al., (2005), most studies have not tried to relate the problem attributes directly to academic achievement.

Therefore, this study aimed to answer the following questions: Which problem characteristics do students and tutors consider generally as contributive to the overall effectiveness of problems in PBL? Do the students' and tutors' perceptions of problem characteristics 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 characteristics, reflect itself in the students' academic achievement? To test this, we conducted focus group interviews with students and tutors separately. The focus group studies were carried out in two phases. In the first phase of group-discussion, we sought the students' and tutors' generalized opinions about characteristics of effective PBL problems. In the subsequent phase, we gathered the students' and tutors' individual responses regarding the effectiveness of eight familiar sample problems. Analysis of the data, results and conclusion are presented in Chapter 3.

**STUDY 3**

Study 3 explored whether it is possible to develop and validate a rating scale instrument derived form the characteristics of problems identified in the earlier studies to measure the quality of problems. Generally two approaches are used to assess problem quality in PBL literature. One approach is to compare the student-generated learning issues with the faculty-intended learning issues (e.g., Dolmans et al., 1993, 1995). Though useful, this approach sheds light on only one problem characteristics; the extent to which the problem leads to the intended learning issues. Furthermore, this approach is time-consuming and resource intensive. Therefore, this approach may not be suitable when evaluating several problems. The second approach which addresses these issues involves the use of a questionnaire. Although few questionnaires are reported in the

PBL literature to assess (1) complexity and structuredness of the problems (Jacobs et al., 2003), (2) problem familiarity (Soppe et al., 2005), and (3) the extent to which problem leads to appropriate learning activities (Marin-Campos et al., 2004), most of these questionnaires are not validated. In addition, only few characteristics are assessed. Therefore the objective of the third study was to develop and validate a more comprehensive problem quality rating scale. To this end, this study first piloted a rating scale before refining to measure five problem characteristics. Chapter 4 of this dissertation reports the validity and reliability testing of the rating scale.

**STUDY 4**

Study 4 attempted to rigorously test the reliability of the rating scale developed in Study 3. Although the earlier study included reliability testing, only internal consistency of the measures was assessed. According to Cohen, Manion, and Morrison (2000) the three principles of reliability are internal consistency, stability, and equivalence. Internal consistency measure of the rating scale indicates the extent of congruency and consistency between the different items of a single scale. On the other hand, stability refers to the measure of consistency over time and across different raters. It is possible that the rating scale is reliable in terms of internal consistency, but only when measured at a certain time-point or when administered to a specific group of students. For instance, high achievers may evaluate the problems differently from low achievers. Hence the objective of Study 4 was to investigate the inter-rater reliability and the stability of the ratings over time, and across different groups of students in using the problem quality rating scale. Chapter 5 presents the findings from Study 4.

**STUDY 5**

Study 5 attempts to extend on Soppe et al.'s (2005) work using the rating scale developed and tested in this research (Studies 3 & 4). In 2005, Soppe et al. carried out an experimental study to investigate the influence of problem familiarity on various aspects of students' learning. Their hypothesis was that students working with the more familiar problem would activate more prior knowledge during the initial discussion. This activation of prior knowledge would stimulate more interest, which would in turn lead to students spending more time on self-study, resulting in acquisition of higher-quality subject matter which is reflected as higher scores on relevant knowledge tests.

To test their hypothesis, they presented one group of students with a familiar version of a problem and another group with an unfamiliar version of the same problem and measured the students' perceptions about problem familiarity, problem quality (interestingness, difficulty level and match with prior knowledge), time spent on self-study using a rating scale. Additionally, external measures such as tutor's assessment of the quality and quantity of learning issues as well as students' achievement in knowledge tests were used. Their results suggested that students found the familiar problem to be of higher quality and interesting than the unfamiliar problem. Surprisingly, no corresponding difference was found in terms of the quality and quantity of learning issues generated by students, time taken for self-study, and students' achievement in knowledge tests as a result of the working on the two problems.

The authors suggested that insignificant differences in some of the measures could be due to the subtleness of the experimental manipulation, and recommended increasing the difference between the familiar and unfamiliar version of the problem. To extend on this work, Study 5 was carried out with some modifications in the methodology. First, two different problems differing in the familiarity level were used instead of using two versions of the same problem. Second, a validated and reliability-tested rating scale (from Studies 3 & 4) was used to measure the various problem characteristics. Third, the tutor's observation of the students' behaviour and learning was included. Chapter 6 presents the findings from Study 5.

Overall, Studies 1 and 2 were explorative in nature and attempted to identify characteristics of problems in students' and tutors' perceptions. Studies 3 and 4 tested the validity and reliability of these characterises while Study 5 attempted to apply the rating scale in investigating the influence of problem familiarity on students' learning. Taken together, these five studies not only tried to shed light on the different characteristics associated with problem quality (Studies 1, 2, 3, 4, & 5) but they also attempted to relate the characteristics to students' learning (Studies 2 & 5). Findings and discussion of the findings for each of the five studies are presented individually in the next five chapters. Finally, to give a more holistic picture, an overall summary and discussion of the five studies are reported in Chapter 7 of the dissertation.

# CHAPTER 2

# CHARACTERISTICS OF PROBLEMS FOR PROBLEM-BASED LEARNING: THE STUDENTS' PERSPECTIVE


## ABSTRACT

This study aimed to capture a 'panoramic' view of problem characteristics that are critical to students in problem-based curriculum. To this end, short essays from biomedical students ($N$ = 34) on characteristics of good problems were text analyzed. Eleven characteristics were identified by students, of which they found the extent to which problem led to the intended learning issues as the most important. Other characteristics included interest triggered by the problem, format of the problem, critical reasoning stimulated by the problem, promotion of self-directed learning, clarity of the problem, difficulty level of the problem, whether a problem enabled application or use, and whether the problem related to prior knowledge, in a decreasing order of importance. Ability of the problem to stimulate elaboration and promote team effort was considered the least important amongst these. These eleven characteristics were clustered into two categories as "features" or "functions" based on the role of the characteristics. Implication and limitations for problem designing and problem evaluation are further discussed.


***Keywords:*** *Characteristics of problems, problems, problem-based learning, students' perceptions*

## INTRODUCTION

Problems are suggested to be one of the three key elements of Problem-based Learning (PBL); the other elements are students and tutors (Majoor, Schmidt, Snellen-Balendong, Moust, & Stalenhoef-Halling, 1990). Problems in PBL refer to the instructional materials presented to students to trigger their learning process and are often formulated in textual format, sometimes with pictures and computer simulations. They are also referred to as “triggers”, “cases”, or “scenarios” in the PBL literature. Problems typically describe a set of situations or phenomena set in real-life context and require the students to explain or resolve (Hmelo-Silver, 2004).

Problems serve to start the learning process. To address the problem, students work in their groups to first discuss and analyze it. This leads to generation of several issues or topics that require further exploration. They then use these unresolved issues or topics as guidelines for their self-directed learning activities. During the period of self-directed learning, they find out more information to answer the problem. Following that, they reconvene, present to one another, and compile the information gathered. This results in integration of their new knowledge in the context of the problem (Hmelo-Silver, 2004). This process of PBL is also referred to as the seven-step model (Schmidt, 1983).

As problems initiate the learning process in PBL, the quality of problem has been suggested to be crucial for students’ learning. To investigate this and examine the interrelationship between the various elements of PBL, Gijselaers and Schmidt (1990) asked students in a PBL curriculum to rate the (1) quality of problems, (2) tutor’s performance, (3) their prior knowledge, (4) the extent of their group functioning, (5) time spent on individual study, and (6) their interest in subject matter using a rating scale. They then analyzed the influence of these key elements on students’ academic achievements by means of causal modeling. In their causal model, they categorized the quality of problems, tutor's performance, and students' prior knowledge as "input" elements; group function and self-study time as "process" elements, and interest and academic achievements as "output" elements. The results showed that of the three "input" elements, the quality of problems had a more direct and stronger influence on the various "process" and "outcome" elements than the other two "input" elements. What this means is that a good problem leads to improved learning. An implication of this result is that learning can be positively influenced by designing better problems. Another study by Van Berkel and Schmidt (2000) confirmed and added support to these findings.

Probing further, Van den Hurk, Wolfhagen, Dolmans, and Van der Vleuten (1999) examined the influence of the quality of problems and tutorial group processes (e.g., breadth and depth of discussion in the tutorial group) on generation of useful learning issues. They found that the quality of the problems indeed had an influence on the generation of useful learning issues. Their findings are supported by an earlier work by Dolmans, Schmidt, and Gijselaers (1995). Dolmans et al. (1995) showed that the overlap between student-generated learning issues and faculty-intended learning issues can be used to assess the effectiveness of a problem. Given the evidence for the importance of the quality of problems, these authors contended that additional information about the nature of the problem is required to improve the quality of problems. They proposed that understanding the characteristics defining the quality of a problem will provide insights to designing and assessing problems in PBL using a rational approach. However, there are not many studies that shed light on the quality of problems in PBL (Jonassen & Hung, 2008).

Generally, problems are designed based on guidelines derived from experiential knowledge, and theoretical principles of learning and cognition (Dolmans, Snellen-Balendong, Wolfhagen, & Van der Vleuten, 1997). For instance, Shaw's (1976) guidelines proposed five dimensions of problems, namely difficulty, solution multiplicity, intrinsic interest, cooperation requirements, and familiarity. Dolmans et al. (1997) outline seven principles of problem design. Their principles are that problems should simulate real-life, lead to elaboration, encourage integration of knowledge, encourage self-directed learning, fit in with students' prior knowledge, interest the students, be of an adequate level of complexity as well as structuredness, and reflect the faculty's objectives.

Hung (2006) proposes a conceptual framework for problem designing in the form a theoretical 3C3R model. The 3C3R model represents the three core components and three process components of problems. The core components refer to "content", "context", and "connection", which underpin the students' content and conceptual learning. On the other hand, the process components; "researching", "reasoning", and "reflecting" represent the students' cognitive processes and problem solving Skills. Jonassen and Hung (2008) focus on one of these problem characteristics – problem difficulty and define it to be characterized by problem complexity and problem structuredness. According to these authors, problem complexity refers to the breadth, attainment level, intricacy, and interrelatedness of problem space while complexity of problem represents the intransparency, heterogenicity of interpretations, interdisciplinary, and dynamicity of problems. Although these guidelines and principles are useful to gain a better understanding about problem characteristics, these are theory-based (Jacobs, Dolmans, Wolfhagen, & Scherpbier, 2003); there is still a lack of empirical studies to validate these theoretical ideas (Jonassen & Hung, 2008).

The few existing empirical studies on characteristics of problems in PBL tend to focus mostly on few specific problem characteristics. For instance, Jacobs, Dolmans, Wolfhagen, and Scherpbier (2003) developed and validated a questionnaire to assess the degree of complexity and structuredness of PBL problems. They defined "complexity" as the number of characteristics or variables that play a role in challenging the students to think and to learn, the interrelationship between these characteristics, and the stability of this interrelationship over time. "Structuredness" of a problem is defined as requiring the application of a limited number of well-structured rules, with solutions that are straightforward and predictable. Therefore, a well-structured problem is thought to have one defined solution compared to an ill-structured problem which may have many possible solutions. They found that although students could clearly differentiate between simple and difficult problems, they had difficulty in discerning ill-structured from complex problems. Hence, the authors classified both ill-structured and complex problem as one factor of problem "difficulty".

Using an experimental approach, Soppe, Schmidt, and Bruysten (2005) investigated the "familiarity" level of problems as a possible characteristic influencing students' learning. They defined familiarity as the extent to which the students can relate to the characters/actors represented in the problem. Their hypothesis was that the familiar version of the problem would activate more of the students' prior knowledge which would in turn stimulate more of their interest, resulting in longer time spent on self-study, and higher achievement scores in knowledge tests. To verify this, they presented two groups of students with either a "familiar" or an "unfamiliar" version of the same problem. The familiar version of the problem was set in a context involving students and their housing facility, while the unfamiliar version used a context of a consultancy firm. The intended learning issues for both the problems remained the same, and pertained to human judgment and decision-making. To measure the influence of the problem, the students were asked to rate the problem they had worked on based on its level of interestingness and familiarity. In addition, other indicators of learning such as the number of explanations generated by students, the quality of the learning issues, amount of self-study time, and amount of knowledge acquired were measured. The results showed that the students perceived the "familiar" version of the problem to be more familiar and interesting than the "unfamiliar" version. However, there was no significant difference in their academic achievements. One possible explanation given was that the difference between the familiar and unfamiliar situation was too subtle, hence resulting in negligible differences in the learning outcomes of the two types of problems. Another possible reason not mentioned by the authors is that although "familiarity" may be one of the meaningful characteristics for rating problems, it might not be the only characteristic. Despite the use of these studies in unearthing more information about the specific characteristics studied (Jacobs et al., 2003; Soppe et al., 2005), a drawback is that they are limited to few characteristics; they do not shed light on the other problem characteristics.

To identify a more comprehensive list of essential problem characteristics, Des Marchais (1999) used a Delphi technique whereby he asked six experts to identify three criteria considered most essential for the design of problems. This Delphi approach led to the identification of nine criteria that were ranked by the experts according to importance. The two most important criteria identified were that the problem should be able to stimulate thinking/reasoning and lead to self-directed learning in the students. (For the other seven criteria see Table 4). Although this study was the first to identify a comprehensive list of problem characteristics using an empirical approach, a point to note is that this study is based on expert's perceptions. It is possible that experts do not experience the problem in the same manner as students. Studies show that students' and tutors' perceptions of various aspects of students' learning may differ (e.g., Zanolli, Boshuizen, & De Grave, 2003). Given that students are the end-users of the problems, it is reasonable to infer that the approach of identifying problem characteristics based on students' experiences is likely to provide a more valuable insight on what types of problems work well. This may raise the question of whether students' perceptions are reliable. Studies have shown that students' rating of instructional context such as teaching skill or the adequacy of instructional materials are reliable and valid (Cohen, 1981).

To include the students' perspective, Schmidt (1985) developed a 59-item rating scale on various aspects of problems and administered to 102 students. The data collated were factor analyzed. A total of eight independent characteristics of problems were identified using this approach. The identified characteristics of problem were learning output, goal clarity, openness, concreteness, familiarity, prior knowledge involved, time on task, and intrinsic interest. Although this study had included the students' perspective, the items of the rating scale were derived based on *a priori* theoretical considerations. That is, the students are restricted to responding on the given characteristics. There is a possibility that students may consider characteristics other than those represented by the rating scale. Hence, we felt that a "bottom-up" approach to understand the students' perspectives is necessary.

In summary, some of the shortcomings of the existing literature on problem characteristics are that (1) they are generally theory-based and not evidence-based, (2) the relatively few empirical literature focus on only few specific characteristics, and (3) studies that have attempted to explore the quality of problems at a broader level are restricted to expert's perceptions or *a priori* theoretical considerations. To address these shortcomings, the present study aimed to investigate the students' perceptions of characteristics associated with good problems using an explorative approach. The questions asked in this study are (1) which are the salient characteristics of PBL problems in students' perspectives?, (2) is it possible to rank the characteristics identified in order of importance?, and (3) what are the implications of the findings? To this end, we asked students to reflect and record their perceptions of what makes a good problem in their e-journals.

## METHOD

### *Participants*

This study was conducted during the second semester of academic year 2006/2007 at the School of Applied Science, Republic Polytechnic in Singapore. A total of 34 second-year students taking "Microbiology" module as part of their course to "Diploma in Biomedical Sciences" participated in this study. Student participants are referred to as B1 to B34.

### *Educational context*

PBL is implemented at Republic Polytechnic in a unique "one day, one problem" approach. The second-year students pursue modules of specialty courses based on their choice of diploma track. Each module comprises of 16 problems and students are required to complete one problem a day. Each day is divided into three meetings with a self-study breakout period between each meeting. The students are presented with the problem in the first meeting. During this meeting, the general outline of the problem is discussed. Students are then given an hour of self-study time to explore further on what they know, do not know, and need to know as well as to gather information. Following this, students and facilitator reconvene at second meeting to discuss on the progress. At the end of this, students are given a second self-study breakout period for 2 hours to compile the information gathered and prepare for their team's presentation. During the third meeting, the students present their findings to the class and discuss. A more detailed description of the PBL process at Republic Polytechnic can be found in Alwis and O'Grady (2002).

### *Procedure*

Students at Republic Polytechnic are required to reflect on the different aspects of their learning process as part of the daily PBL sessions and record their reflections in personal online journals. Participants in the study were asked to write a short essay as part of their reflection on what they considered to be characteristics of good problems. The question administered to the participants was "What is your perception of a good problem trigger and why? You can base your answer on any of the problems you have done so far."

### *Analysis*

The essays written by the participants were compiled and analyzed using TextSTAT text analysis software, obtained from the web link http://www.niederlandistik.fu-berlin.de/textstat/ (Huning, 2007). TextSTAT is a simple concordance program for the analysis of texts using data in the ASCII/ANSI/ HTML/ Microsoft Office format. The program is designed to count the word frequency in the input data. The assumption was that the more often a characteristic was mentioned, the more important it was for students. The example in Table 1, gives an impression of how the software works. It contains the complete response by participant B5.

**Table 1**

*Complete Essay Response by Participant B5*

| |
|---|
| Personally I feel that a good problem trigger should be something interesting yet easy for us to understand. I would like it to be interesting so that I would not get bored while researching for the information. Besides that, it would be good for it to be slightly difficult as what is the use of a problem trigger if it doesn't trigger the mind and make us think out of the box. I don't prefer problem triggers that are too easy and straightforward because it just seems too easy to be true and we might finish our task too fast. Thus not making full use of the time given from the 2nd breakout till the 3rd meeting. Nevertheless, I do not prefer them to be too difficult because at times the topic that we need to touch on is quite a lot yet there is not much time to research and comprehend the findings before presenting. There was a trigger which I think is interesting and all of the above. It was a problem trigger from one of the biochemistry lessons (last semester). The problem trigger was in a form of a riddle. To me it was fun and interesting, as we need to crack our head to solve and understand the problem trigger. It goes like this:<br>"Thin or brawn<br>Men flex them with valor<br>Women have it permed and straightened,<br>For more than a dollar<br>Acrylic is out<br>Manicures are in<br>All the above<br>Are made of the same thing<br>Some soft<br>Others hard like pine<br>Take away their differences<br>What will you find?"<br>I feel that if problem triggers would be interesting it would give us the drive to do work/research. Furthermore if it is difficult to a certain extent, it will enable us/me to think hard and at the same time have a better discussion within the team and class. |

The word document of the excerpt in Table 1 was entered as the input file to be processed by the text STAT software. The program generated a frequency list, with a total of 319 words, sorted out in descending order. A portion of the top-most frequency is shown in Table 2.

**Table 2**

*Partial Frequency List in Descending Order as Generated by the textSTAT Program*

| Word* | Frequency | Word | Frequency | Word | Frequency |
|---|---|---|---|---|---|
| the | 17 | that | 6 | difficult | 3 |
| to | 12 | is | 6 | time | 3 |
| it | 10 | would | 5 | think | 3 |
| and | 10 | interesting | 5 | for | 3 |
| a | 10 | too | 4 | easy | 3 |
| trigger | 7 | was | 4 | like | 3 |
| be | 7 | us | 4 | feel | 2 |
| problem | 7 | not | 4 | prefer | 2 |
| of | 7 | if | 3 | because | 2 |
| I | 7 | we | 3 | our | 2 |

** Text document of participant B5's response was used as the input file*

From the list generated, appropriate evaluative words associated with various qualitative aspects of problems were identified manually. For example, in the given frequency list in Table 2, only 5 words, namely, “interesting”, “difficult”, “time”, “think” and “easy” were identified to be appropriate. The filtered words were then categorized based on semantic similarity. A few examples of words used by participants for each aspect are provided in Table 3. The frequency percentage of selected words in the different categories was computed to rank the identified characteristics.

**RESULTS**

The students who participated in this study were all in the second semester of second year in the PBL curriculum. On average, each student had worked on over a hundred problems. Students had each taken 4 modules per semester, consisting of 16 problems per module. Drawing on the participants’ experience of solving this many PBL problems was considered to be useful in providing insights on the problems. Essay responses from the 34 participants totaling 6580 words were frequency counted. Of this, only 994 words were deemed appropriate as describing the problem characteristics. These words were then categorized into eleven characteristics based on semantic similarity and ranked based on word frequency percentage, as summarized in Table 3.

The most important characteristic of the problem in students’ view was that the problem should lead to the intended learning issues, which accounted for approximately 24% of the responses. The least important characteristic in students’ view was problem promoting teamwork, with a frequency percentage of only 2%. Samples of student responses pertaining to each characteristic are presented in Appendix A.

**Table 3**

*Key Characteristics of Problems and Ranking by Importance to Students*

| A problem should… | Words used by students | Frequency percentage of words used | Ranking of importance * |
|---|---|---|---|
| lead to learning issues | learn, issues, facts | 23.8 | 1 |
| trigger interest | interesting, like, capture | 11.5 | 2 |
| be of suitable format | phrase, picture, sentence | 10.9 | 3 |
| stimulate critical reasoning | thoughts, ideas, logic | 10.2 | 4 |
| promote self-directed learning | research, explore, tackle | 10.0 | 5 |
| be of suitable clarity | obvious, clear, understand | 7.3 | 6 |
| be of appropriate difficulty | easy, difficult, hard | 7.1 | 7 |
| enable application or use | apply, world, use | 7.0 | 8 |
| relate to prior knowledge | know, remember, background | 6.7 | 9 |
| stimulate elaboration | elaborate, brainstorm, discuss | 3.6 | 10 |
| promote teamwork | team, class, together | 1.9 | 11 |

*According to scale of importance from 1 to 11, 1 being the most important

**DISCUSSION**

This study aimed to explore the students' perspectives on the characteristics of good problems in PBL based on their experiences with problems. The specific questions asked were (1) which are the salient characteristics of PBL problems in students' perspectives?, (2) is it possible to rank the characteristics identified in order of importance?, and (3) what are the implications of the findings? The driving force behind the study was the lack of empirical studies which look at the wider spectrum of problem characteristics, especially from students' perspectives. The underpinning supposition of this study is that students' perspectives of the characteristics are useful since they are the end-users of the problems.

To this end, 34 first-year students from PBL curriculum were asked to reflect in their e-journals on what they deemed as characteristics of good problems. Their responses were text analyzed based on semantic similarities resulting in the identification of eleven characteristics. These were then ranked according to their frequency counts. The results show that the students found (1) the extent to which the problem leads to the intended learning issues as the most important characteristic. Other characteristics identified were (2) interest triggered by the problem, (3) format of the problem, (4) the extent to which the problem stimulated critical reasoning, (5) the extent to which the problem promoted self-directed learning, (6) clarity level of the problem, (7) difficulty level of the problem, (8) the extent to which the problem is relevant; that is applicable and useful, and (9) the extent to which the problem relates to the students' prior knowledge, in a decreasing order of importance. (10) The extent to which the problem stimulates elaboration and (11) promotes team effort were considered the least important amongst these. Overall, the results indicate that it is possible to identify a wider spectrum of problem characteristics based on the students' perspectives.

A comparison of the eleven characteristics from this study with the literature on problem characteristics (Des Marchais, 1999; Dolmans et al., 1997; Shaw, 1976) shows that the students also identified similar characteristics as those proposed in the literature. See Table 4 for the comparison of students' perspective from this study with other empirical studies (e.g., Des Marchais, 1999), and theoretical guidelines (e.g., Dolmans et al., 1997). This could be possibly because the students are constantly exposed to the views of constructivist learning as part of their PBL curricula. Hence they may align their beliefs with the principles of constructivist learning that learning occurs as a result of engaging in self-directed learning as well as collaborative work to find solutions to authentic problems, which results in gain in their content knowledge, and interest (Savery & Duffy, 1995). As to whether students associate these principles in practice, Loyens, Rikers, and Schmidt (2007) showed that students do recognize these constructivist assumptions.

**Table 4**

*Summary of the Characteristics of Problems for Problem-based Learning Stipulated by Various Studies*

| 11 criteria<br>This study<br><br>(Ranked by importance indicated by the prefix)<br><br>Problem should… | 8 criteria<br>Schmidt (1985)<br><br>(No ranking) | 9 criteria<br>(Des Marchais, 1999)<br><br>(Ranked by importance indicated by the prefix)<br><br>Problem should… | 7 principles<br>Dolmans et al. (1993)<br><br>(No ranking) | 5 features<br>Shaw (1976)<br><br>(No ranking) |
|---|---|---|---|---|
| 1. Lead to learning issues | Learning output<br><br>Openness | 5. Lead to discovery of learning objectives | Should match one or more of the faculty objectives | Solutions multiplicity |
| 2. Trigger interest | Intrinsic Interest | 6. Arouse curiosity and interest | Enhance students' interests | Intrinsic interest |
| 3. Be of suitable format | | | | |
| 4. Stimulate critical reasoning | | 1. Stimulate thinking, analysis and reasoning | | |
| 5. Promote self-directed learning | Time on task | 2. Initiate self-directed learning - | Stimulate self-directed learning | |

| | | | | |
|---|---|---|---|---|
| 6. Be of suitable clarity | Concreteness<br><br>Goal clarity | 9. Contain appropriate medical analytical vocabulary | | |
| 7. Be of appropriate difficulty | | | | Difficulty |
| 8. Enable application or use | | 7. Be on topics related to public health<br><br>8. Consider a global perspective | In context that is relevant to future work | |
| 9. Relate to prior knowledge | Familiarity / Prior knowledge | 3. Relate to previous basic knowledge<br><br>4. Be proposed in a realistic context / Relate to previous basic knowledge | Adapt well to students' prior knowledge<br><br>Relevant to basic concepts in the context of a clinical | Familiarity |
| 10. Stimulate elaboration | | | Contain several cues to stimulate students to elaborate | |
| 11. Promote teamwork | | | | Cooperation requirements |

Comparing the eleven characteristics from this study with Des Marchais' list of nine characteristics (1999), we can see that the students identified all of the nine problem characteristics cited by the experts. In addition, the students identified new problem characteristics such as problem format, problem difficulty, the extent to which the problem stimulates discussion, and the extent to which the problem promotes teamwork. More noticeably, the students differed from the experts in the ranking of the problem characteristics. For instance, the experts in Des Marchais' study (1999) identified the two most important criteria as (1) the extent to which the problem stimulates thinking/reasoning and (2) the extent to which the problem leads to self-directed learning in the students. However, the students in this study identified the extent to which the problem leads to intended learning issues as the most important characteristic. An explanation for the differences observed could be the different roles of the experts and students. Hence, their expectations of the quality of problems could be different.

In line with this are studies which show discrepancies between the students' and tutors' perceptions of PBL. For instance, Gerzina, McLean, and Fairley (2005) showed that students and tutors differed significantly in their perceptions of the extent to which theoretical knowledge was applied in clinical settings; that is, more students than tutors perceived a link between the theory and application of it in the concerned dental clinical teaching program. In another study, Zanolli, Boshuizen, and De Grave (2002) showed that students and tutors differed in their ratings of several aspects of PBL. While the students ranked the tutors as the most important factor for their learning, the tutors ranked the students as the most import factor for the same. In addition, the students and tutors disagreed significantly on factors such as assessment and problem, with the students generally having a higher means than the tutors.

Another possibility could be that the objective and implementation of PBL curriculum in Rouen University and the Polytechnic in this study are different (Des Marchais, 1999). On the one hand, Des Marchais' (1999) study was conducted in a Medical University, while, the present study was conducted in a Polytechnic which employs PBL across all modules. Schmidt, Van der Molen, Te Winkel, and Wijnen (2009) pointed out that the implementation of PBL varied depending on the objectives of PBL, and has proposed a categorization of the various versions of PBL into three types based on its objectives. Type I PBL focuses on information processing and is founded on the cognitive psychology principles of mental-model construction. Type II PBL is process-oriented, focusing on problem-solving skills such as clinical reasoning, and type III PBL focuses on learning skills which help students learn how to learn. As the PBL curriculum in Des Marchais' study (1999) was situated in a medical context, it is possible that it focused more on problem solving skills, and is of Type II PBL. On the other hand, the institution involved in this study has adopted PBL across its curricula and focused more on knowledge construction as in type I PBL. Although reasonable, this postulation needs to be examined further. One way to overcome this difficulty in future studies will be to compare the perspectives of students and tutors from the same institution/the same type of PBL curricula. Understanding the difference in student-tutor perceptions will be important in interpreting program evaluation by the two groups.

Assuming that the most vital characteristics of problems are likely to be cited by most if not all the various studies, we can conclude from Table 4 that the characteristics of problems to (1) lead to the intended learning issues, (2) trigger interest, and (3) relate to prior knowledge are vital in determining the quality of PBL problems. Cross-referencing this with the students' perceptions that the most important characteristic is the extent to which the problem leads to intended learning issues, it is reasonable to infer that this characteristic is indeed of high importance to the quality of problems. This is not surprising as one of the main objectives of PBL and constructivist learning is that students are able to construct new knowledge in relation to the problem presented (Mayer,1999). In an earlier study Schmidt (1985) also reported that students rated the most important characteristics to be the amount of knowledge gained from working on the problem. In agreement with this, other researchers have attempted to evaluate the effectiveness of problems in PBL by comparing the student-generated learning issues with the faculty-intended learning issues (Dolmans, et al.1993, 1995; Mpofu, Das, Murdoch, & Lanphear, 1997). Their hypothesis is that an effective problem will result in higher congruence between the student-generated learning issues and faculty-intended learning issues. The results suggest that it is indeed possible to evaluate the effectiveness of problems using this characteristic. For instance, by comparing the two sets of learning issues across 12 problems, Dolmans et al. (1993) showed that students identified 64% of the faculty-intended learning issues on average.

A second characteristic cited across the various studies is the interestingness of the problems. This characteristic is reflective of the underpinning principles of constructivist learning that the learning process should trigger students' interest (Mayer, 1999; Savery & Duffy, 1995). Gijselaers and Schmidt (1990), Schmidt & Gijselaers (1990), and Van Berkel and Schmidt (2000) showed that the quality of problem has a positive influence on students' interest and learning. There are also several other studies which showed that group discussion on the problem positively influences students' intrinsic interest in the subject matter (e.g., De Volder et al., 1986). Studies by Soppe et al. (2005) showed that problems found to be familiar triggered more interest. Thus, this characteristic seems to be important in PBL.

The third characteristic cited across the various studies is that the problem should relate to students' prior knowledge. This can also be defined as the "familiarity level of the problem" in terms of both content and context. This problem characteristic relates to the cognitive psychology principles that activation of prior knowledge in a collaborative group is needed to co-construct new knowledge. Several studies also supported the notion that prior knowledge strongly influences learning (Anderson, 1990; Dolmans, Wolfhagen, & Schmidt, 1996; Mamede, Schmidt, & Norman, 2006; Norman & Schmidt, 1992; Schmidt & Boshuizen, 1990; Soppe, et al., 2005).

The characteristic "clarity level of the problem" can be defined as the extent to which the problem is comprehendible and transparent to the students. The students' responses on this characteristic support the idea that this characteristic has a close association with the extent to which the problem leads to the intended learning issues. For instance, one of the students elaborated that "A good problem must contain clue words of the topic being taught for the day. Even if it is without any help of the worksheet, at least we know what we had to learn" (Participant B11). In line with this, Mayer (1999) proposes that techniques such as using headings, providing a summary of information or including additional questions and statements in the problem design can help students towards the important learning issues. Specification of goals in the problem can be a means to influence the level of problem clarity. A study by Verkoeijen, Rikers, Te Winkel, and Van der Hurk, (2006) showed that goal specificity results in generation of more quality and quantity of learning issues in the discussion phase than a goal-free problem. However, the goal-free problem was shown to have led to more quality and quantity of learning issues in the reporting phase than the goal-specified problem. In addition, this study revealed that the goal-free problem had a positive influence on the study time, number of articles read and time used for the reporting phase. Although these studies shed light that the clarity level of the problem influences students' learning, it is not yet clear as to how clear the problem should be. Hence further research on this characteristic is still needed.

The characteristic "format of the problem" can be defined as the physical representation of the problem as in whether it is in textual format, if it includes an illustration, and whether the problem is short or lengthy. Students' response on this characteristic indicates that the format of the problem has an influence on the interest triggered. For instance, one student wrote that "My definition of good problem is firstly, it has to be straightforward, and no to long-winded ones. It is the start of morning, a good problem can trigger off enthusiasm, if it is long-winded, honestly, it can kill off the learning spirit" (Participant B5). The cognitive load theory suggests that format of the instructional material may influence the learning efficacy of a learning environment (Hoffler, & Leutner, 2007). As such, it will be valuable to find out using empirical approaches on how the format of the problem can be utilized to engage the students.

Next, students mentioned that the problem should be of appropriate difficulty level. From their response, it could be inferred that difficult problems may not be all that bad. One participant noted that "It would be good for the problem to be slightly difficult as what is the use of a problem if it doesn't trigger the mind and make us think out of the box. I don't prefer problem that are too easy and straightforward because it just seems too easy to be true and we might finish our task too fast. Thus, not making full use of the time given from the 2nd breakout till the 3rd meeting. Nevertheless, I do not prefer them to be too difficult because at times the topic that we need to touch on is quite a lot yet there is not much time to research and comprehend the findings before presenting" (Participant B5).

The concept of the difficulty level of problems has been discussed elsewhere in PBL literature. For example, Jonassen and Hung (2008) defined problem difficulty as problem complexity and structuredness. They further elaborated that problem complexity refers to the breadth, attainment level, intricacy, and interrelatedness of problem space while complexity of problem refers to the intransparency, heterogenicity of interpretations, interdisciplinary and dynamicity of problems. Jacobs et al. (2003) attempted to investigate whether students can conceptually distinguish between problem complexity and structuredness using a questionnaire. They found that although students can distinguish between too simple and too structured problems, they were not able to discern ill-structured and complex problems. Therefore they combined these too factors and classified the new factor as "problem difficulty". Taken together, these results show that despite the importance of problem difficulty in problem designing, it remains elusive, and that further research is needed on this characteristic of the problem.

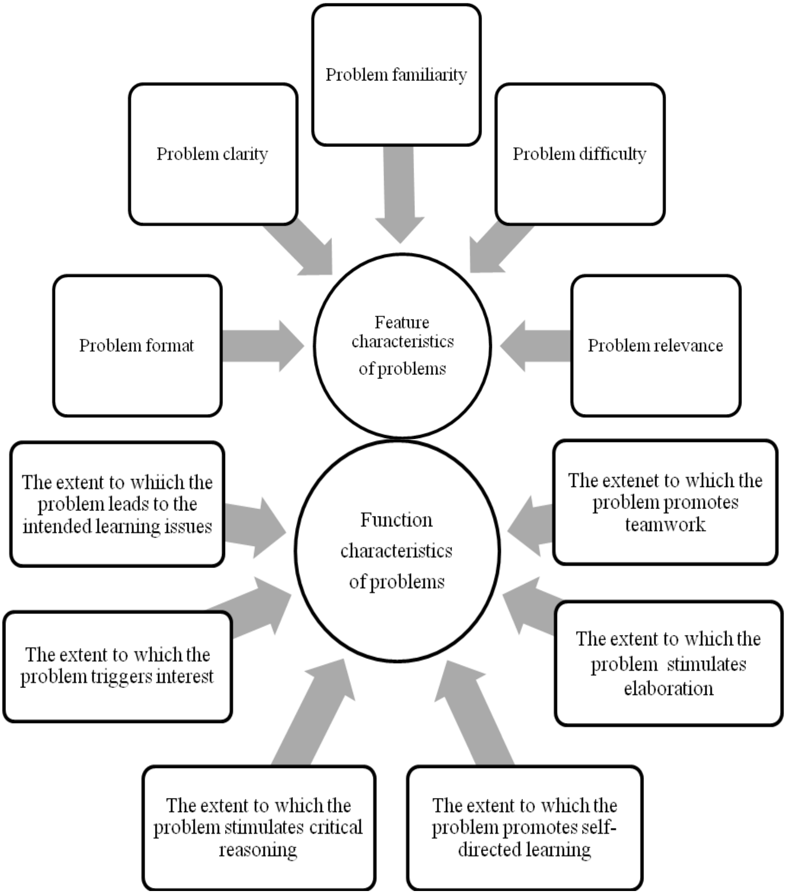
Another characteristic is the extent to which the problem is perceived as relevant that is, applicable or useful. PBL is founded on the principle that students not only acquire knowledge but that they know how to apply this knowledge in different situations that represent the real-world. Thus, use of authentic contexts is recommended to be used in PBL (Savery & Duffy, 1995). Research on learning shows that information learned in context is better recalled and retained (Brown, Collins, & Duguid, 1989). In addition, problems that are perceived to be relevant are also likely to engage the students in the learning process and contribute to their learning. For instance, Araz and Sungur (2007) showed that task value, was one of the factors which had both direct and indirect effects on achievement. The other factors were reasoning ability, learning approach, prior knowledge and motivational variables.

Two other characteristics of problem: the extent to which the problem promotes self-directed learning and the extent to which the problem stimulates critical reasoning are also reflective of the constructivist principles (Savery & Duffy, 1995), and focuses on the problem solving skills and learning to learn skills (Mayer, 1999). These characteristics are likely to be more highly regarded especially in institutions which adopt type II and type III PBL (Schmidt et al., 2009); that is PBL curricula in which the main objectives are to teach students problem solving skills and learning to learn skills respectively. PBL has been shown to have an influence on students' critical reasoning (Albanese & Mitchell, 1993) and self-directed learning (Blumberg & Michael, 1991). This could be the result of the influence of several variables such as quality of problems, role of tutors, learning environment. Hence looking at the various variables at the same time may result in confounding results. Therefore it may be useful to look at the variables one at a time. For instance, future research could explore how a certain characteristic of problems such as problem difficulty stimulates critical reasoning and promotes self-directed learning.

The students in this study also felt that the problem should stimulate elaboration and that it should promote teamwork. Interestingly, the experts in the Rouen Delphi study had not cited these two characteristics. Again, it is likely to be the result of the different roles played by the students and experts. This stresses the need for further studies to probe both the students' and tutors' perspectives. In a way, this characteristic can be seen to be a reflection of the constructivist learning principle that learning takes place in collaborative work (Savery & Duffy, 1995). A possible explanation for the low importance accorded to these problem characteristics could be that they are associated more with the tutor due to the element of social interactivity.

Overall, this study attempts to present a wider spectrum of the characteristics of good problems in students' perspective. In considering the implication and limitation of the findings, we propose that these eleven characteristics can be classified into two groups as either "features" or "functions", based on their roles. "Features" of the problems refer to characteristics that are design elements of the problems. Characteristics such as problem format, clarity, familiarity, difficulty and relevance (application and use) are such design elements of the problems. On the other hand, "function" characteristics refer to the potential outcomes of engaging with or working on the problems. Of the eleven identified characteristics, the extent to which the problem stimulates critical reasoning, promotes self-directed learning, stimulates elaboration, promotes teamwork, stimulates interest, and leads to the intended learning issues are such functional properties. In a way, these functional characteristics are reflective of the five principles of constructivist learning and the objectives of PBL (Savery & Duffy, 1995, Mayer, 1999). Figure 1 shows the classification of the proposed feature and function characteristics.

**Figure 1**
*Function and Feature Characteristics of Problems in Problem-based Learning*



Hence, further research is needed to unravel this complex relationship to understand how to design effective problems. As for problem evaluation, the results add further support to the existing understanding that the extent to which the problem leads to the intended the learning issues is an important indicator of the problem effectiveness. In addition, the results add other characteristics that need to be considered in evaluating the effectiveness of problem. For instance, the functional characteristics of the problems are likely to serve as appropriate indicators of problem effectiveness as these characteristics represent the objective of PBL. Therefore, measuring these characteristics could be used to indicate to what extent the problem plays a role in the effectiveness of PBL. To evaluate the effectiveness of problems, future studies could investigate how the feature characteristics influence the function characteristics of problems.

Although earlier studies have identified various characteristics of problems, these are not classified further. The exception is Hung's classification of design elements as "core" components and "process" components (Hung, 2006) whereby the core components refer to elements such as content, context and connectivity which support context /concept learning. While the process elements refer to elements as researching, reflecting and reasoning which support cognitive processes of learning and problem-solving skills. We propose a different classification of the characteristic as feature and function in an attempt to identify characteristics that can be manipulated and considered in designing the problem to enhance students' learning. Our proposition is that the feature characteristics can be manipulated to bring about an effect on the function characteristics.

However, there are number of limitations to this study. The first limitation is that the students' essay responses were used to derive the characteristics of the problems for PBL. Hence the study is limited by students' vocabulary spectrum. Moreover, the risk is that if students are not able to recognize the different characteristics of the problems, they will not be able to mention these concepts, and thus increasing the chance of overlooking these characteristics. Nevertheless, results from this study shows that students participating in this study were able to use words that match with characteristics identified by theoreticians and experts. Second, there is a possibility that some words can be categorized into one or more characteristics of problems. For instance, the word "long" could be associated with problem format as well as the extent to which the problem promoted self-directed learning as the word could be in reference to length of time. The implication of this is that the ranking of the importance of characteristics based on frequency count of the words may not be absolute. However, from the students' response, it is inferable that the students do consider the eleven characteristics. This stresses the need for future studies to develop an instrument such as a rating scale to measure the characteristics of problems with higher reliability. Third, the tutor's perspective from the same institution/institution which adopts similar type of PBL (Schmidt et al., 2009) was not included. This presents itself for further work. Fourth the students were not given concrete sample problems to refer to. Hence the participants could have mentally referred to different problems. In future studies, students could be given concrete sample problems to refer to. In conclusion, this study has only taken the first steps in identifying the various characteristics associated with good problems in PBL. Further research is needed to find out how these characteristics interact with each other in influencing the students' learning.

# CHAPTER 3

## STUDENT AND TUTOR PERCEPTIONS ON CHARACTERISTICS OF EFFECTIVE PROBLEMS IN PROBLEM-BASED LEARNING

### ABSTRACT

This study aimed to identify the characteristics that students and tutors associate with effective PBL problems, and assess the extent to which these characteristics actually relate to the effectiveness of problems. To this end, students and tutors were asked to discuss in focus groups about the possible characteristics of effective problems, and individually judge the effectiveness of eight sample problems that they had worked on. Text analysis of the focus group discussion transcripts identified eleven problem characteristics. These characteristics were subsequently used to frequency-score participant judgments of sample problems. Relating tutor and student judgments with student grades yielded high and significant correlations, suggesting that the eleven problem characteristics generated reflect aspects of problem effectiveness.

***Keywords****: Characteristics of problems, Problem-based Learning, 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) number of lectures are limited. The latter is in line with 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 characteristics are potentially of importance. Well-designed problems may, in principle, lead to better learning. Indeed, existing studies demonstrate the impact of problem quality on students' learning. For instance, Gijselaers and Schmidt (1990) investigated how the quality of problems related to 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 the group functioning and time spent on individual study, and through these on interest in subject matter and academic achievement (See also Schmidt & Gijselaers, 1990 and Van Berkel & 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). They are 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 students in problem-solving, to rekindle their prior knowledge, to spark discussions, to encourage collaborative work, to promote self-directed learning skills, and to lead to acquisition of relevant content knowledge in the course of tackling the problem (Barrows & Tamblyn, 1980; Hmelo-Silver, 2004). When a problem is presented to students at the start of the small-group session to trigger the learning process, the students confront the problem using their own prior knowledge and knowledge offered by their teammates. The 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 (Schmidt, 1983). Overall, the learning process in PBL is self-directed by students and is more problem-centered than teacher centered.

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. PBL tutors 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, Mpofu, Hasan, & Stewart, 2002; Maudsley, 1999). Despite being actively engaged in the students' learning process, the PBL tutors do not teach the content knowledge directly to the students as done by teachers in conventional curriculum. Instead, the content knowledge is synthesized by the students based on their self-study and discussions, which is in turn guided by the nature of problems. This role transition of tutors and students in PBL emphasizes the importance of problems in the learning process.

If the quality of problems makes a difference in terms of student learning, then questions can be raised about designing effective problems. Discerning characteristics of effective problems is expected to provide insight into designing better problems, refining existing problems, and evaluating the quality of problems. Furthermore, it could improve our current understanding on how problem quality influences the learning process and outcomes.

Traditionally, guidelines based on cognitive theories and experiential knowledge are used to design problems. Notably, Shaw's five problem characteristics (1976), Dolmans' seven principles of case design (1997), and Woei Hung's conceptual framework for designing PBL problems (2006) have provided theoretical dimensions of problems. However, only limited empirical studies that describe the problem characteristics exist. By and large, most of the empirical studies have referred to few (one to three) problem characteristics and are usually based on either students' or tutors' opinions, but not both.

Dolmans, Schmidt, and Gijselaers (1995) investigated the effectiveness problems by comparing the learning issues generated by students with those intended by faculty for twelve problems. The idea behind the comparison was that effective problems will lead students to the intended learning issues, and in this case, there should be a match between the student-generated and faculty-intended learning issues. From the student responses, they found that an average of 64% of the intended learning issues across the twelve problems was identified by the students. As for the gaps in identifying some of the faculty-intended learning issues, the students attributed it to the complexity and unfamiliarity levels of the problems.

Jacobs, Dolmans, Wolfhagen, and Scherpbier (2003) validated a questionnaire intended to assess the degree of complexity and structuredness of PBL problems based on students' responses. 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, while, ill-structured problems have many possible solutions. Their results suggested that students considered problem structuredness to be more important in determining problem quality.

Soppe, Schmidt, and Bruysten (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 and were kept unaware of this manipulation. Students' self-report on the problem characteristic 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.

To classify a wider spectrum of problem characteristics, a recent review by Kim et al. (2006) explored one hundred studies from various disciplines. A total of five problem characteristics were delineated. They are that problem should be relevant, realistic, engaging, challenging and instructional. 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 characteristics. These characteristics 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 issues, (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 characteristics.

Des Marchais' study (1999) was the first to identify a broader spectrum of problem characteristics. However it had two limitations. The first limitation is that the expert responses were generalized opinions about PBL problems and 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. The second limitation is that only expert judgments were considered in their study. As students are the end-users of the problem, investigation of the 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 characteristics focus on few characteristics and do not incorporate both students' and tutors' perceptions. In addition, asking participants to mention desirable characteristics of problems in general may yield different answers as compared with asking them to mention characteristics of specific problems. Finally, with the exception of Soppe et al. (2005), most studies did not try to relate problem characteristics directly to academic achievement. Therefore, the present study aimed to answer the following questions: which problem characteristics do students and tutors consider generally as contributive to the overall effectiveness of problems in PBL?, do the students' and tutors' perceptions of problem characteristics hold across a set of problems?, to what extent do students and tutors agree in their judgments of the overall effectiveness of these problems?, and does the evaluation of problem effectiveness based on the identified characteristics reflect itself in the students' academic achievement?

To address the above questions, we conducted focus group interviews with eleven students and five tutors. Each focus group consisted of two to three students or tutors, and the interviews were conducted in two phases. In the first phase of group discussion, we sought the students' and tutors' generalized opinions about characteristics of effective PBL problems. In the subsequent phase, we gathered the students' and tutors' individual responses regarding the effectiveness of eight familiar sample problems. Transcripts of the discussions from the first phase were text analyzed to identify the characteristics of effective problems in general. These characteristics were then used as criteria to frequency-score the students' and tutors' individual responses about the familiar problems. The resulting scores were used to compare the students' and tutors' perceptions of problem characteristics associated with the effectiveness of the sample problems and to relate their perceptions with actual student grades for the subject matter covered by each problem.

## 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. For the second phase of the study, student achievement data from 2566 students were used.

### *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 overall 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 textual format; with P4 being the longest problem at two and a half A4 pages length. The other problems were shorter than one A4 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 Appendix B 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 & 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 study was 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 minutes 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 effective problem to you, based on your experience?" The focus group discussions were audio-recorded and transcribed for further analysis.

Following the discussion, the 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 of each problem?" They were also informed that *positive aspects* of the problem refer to the problem characteristics deemed as contributive to the overall effectiveness of the problem whilst *negative aspects* refer to characteristics 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 characteristics) 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 notepads, 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 on to the next table until they had visited all eight tables. The setup was as such that the participants had no opportunity to talk to, 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 characteristics. The assumption was that the more often a characteristic-associated word was mentioned, the more important the characteristic 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 characteristics. These words were then classified based on semantic similarity into three problem characteristics: *problem interestingness* (like, motivating, challenged, enthusiastically), *problem promoting self-directed learning* (solving, doing), and *problem leading to learning issue* (learning, objectives).

To analyze the data from the second phase of the study, the newly identified problem characteristics 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 reasoning* (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.

To investigate if the identified problem characteristics influenced the effectiveness of problems, the tutor judgments represented as the *observed overall tutor score* for the eight problems were correlated with the average daily-grades obtained by the entire cohort of 2,566 students taking the module for the respective problems. For the analysis, Pearson's correlation coefficient was used.

## RESULTS

Text analysis of the discussion transcripts revealed that students and tutors associated a total of eleven problem characteristics with effective PBL problems in general. The identified characteristics 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 issues. However, as Table 1 shows, there were some differences between students and tutors in ranking the remaining characteristics.

**Table 1**

*Key Characteristics of Effective Problems According to Students and Tutors*

| Students* An effective problem should… | Tutors* An effective problem should… |
|---|---|
| lead to appropriate learning issues (37.4%) | lead to appropriate learning issues (25.1%) |
| promote self-directed learning (22.1%) | promote self-directed learning (24.6%) |
| stimulate critical reasoning (14.5%) | trigger interest (16.0%) |
| promote teamwork(10.7%) | be of suitable format (8.2%) |
| trigger interest (4.6%) | stimulate critical reasoning (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 characteristics were ranked according to frequency of connoting words

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 characteristics. In other words, the eleven general problem characteristics were held in considering the effectiveness of problems at the micro-level of specific problems. Hence we used these characteristics 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.

**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 |

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, $\chi^2$ (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, $\chi^2$ (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, $\chi^2$ (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. On the whole, P4 was considered to be the least effective problem while P2 was considered to be one of the most effective problem by both the students and tutors. A copy of the problem P2 and problem P4 is attached in the Appendix B for reference.

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 students 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 characteristics associated with effective PBL problems. Second, verify if these generalized perceptions are held in judging the effectiveness of specific problems. Third, we wanted 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 characteristics actually relate to the problem effectiveness. To this end, a focus group approach consisting of two phases; a group discussion phase and an individual response phase was undertaken. The group discussion phase gathered students' and tutors' generalized opinions about the characteristics that make the problems effective. Results suggest that both the students and tutors associated a total of eleven characteristics with the effectiveness of PBL problems in general. The subsequent phase of the study was conducted to collect student and tutor responses on specific problems. When the identified characteristics were used as criteria to frequency-score the individual student and tutor responses on the eight sample problems and these scores were related to student grades, three findings were made: (1) the eleven general characteristics of effective problems were considered by the students and tutors in judging specific problems as well, (2) there was no significant difference between the students and tutors in their judgments about the overall effectiveness of the eight sample problems, and (3) 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 characteristics, derived from both student and tutor opinions, turned out to be related to student learning.

In answering the first question, we generated eleven characteristics of effective problems based on focus group discussion with both students and tutors. To see if the eleven characteristics measured the same aspects mentioned in the other studies, the characteristics generated in this study were compared with those cited in the earlier mentioned literature. The eleven identified characteristics of effective problems largely covered the characteristics mentioned by the various other studies, including the nine characteristics mentioned in Des Marchais' study (1999). We found three characteristics to be common amongst the various studies. They were: (1) a problem should lead to formulation of appropriate learning issues, (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 characteristics but not others. Nevertheless, we infer that the commonality of the three characteristics regardless of the differences in the various studies imply importance of these characteristics in designing problems. In addition, this study identifies a unique characteristic not mentioned in the other studies – that is the *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. With regard to instructional design, the cognitive load theory suggests that format of the instructional material may influence the learning efficacy of a learning environment (Hoffler, & Leutner, 2007). As such, an investigation of the influence of problem format on students' learning process and outcomes in PBL would be worth exploring. Next, when we compared whether the students and tutors differed in their generalized perceptions of PBL problems, we found that students and tutors alike emphasized that an effective problem should lead to appropriate learning issues. This characteristic is also considered by Dolmans et al. (1995) and Mpofu, Das, Murdoch, and Lanphear (1997) as an important characteristic of effective problems.

To answer the second question and verify if the generalized characteristics 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 characteristics were generated by the students and tutors when considering specific problems and problem in general. There were no new characteristics generated when referring to specific problems. Implication of this result is that the eleven characteristics may be used to assess the effectiveness of specific problems as well as problems in general. Hence we used the eleven characteristics as frequency-scoring 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 in the students' learning process in PBL, there is frequent communication in the form of feedback from tutors to students and discussion between students and tutor about the learning progression (Schmidt, 1983), the two groups could have noted similar elements influencing 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 characteristic 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 reasoning. 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 issues 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 characteristics 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 characteristics 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 characteristics 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 characteristics 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 characteristics. Assessment of the effectiveness of sample problems using the eleven characteristics 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 characteristics are related to the effectiveness of the problems. Compared with other studies in literature, this study seems to be the first to collate a list of characteristics 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 & Lymn., 2008); but not both. This study also attempts to extend beyond identifying the characteristics by relating the eleven identified characteristics 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 characteristics of PBL problems did not relate the characteristics to the academic achievements.

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 characteristics influences the students' learning. Hence further investigation on this is needed. A second limitation is that this study is retrospective. Utility of the identified problem characteristics as criteria to predict quality of untried problems remains to be investigated. A feasible follow-up study is to develop a problem quality rating scale based on the eleven problem characteristics and validate it to gain a deeper understanding of the role of the eleven problem characteristics in problem effectiveness. This will provide a better insight as to what extent each of the identified characteristics leads to overall problem effectiveness and how these characteristics are inter-related.

# CHAPTER 4

# CONSTRUCT VALIDITY AND RELIABILITY OF A RATING SCALE TO ASSESS THE QUALITY OF PROBLEMS IN PROBLEM-BASED LEARNING

## ABSTRACT

This study aimed to assess the construct validity and reliability of a newly devised problem quality rating scale. The rating scale, consisting of 32-items and measuring five characteristics of problems in Problem-based Learning, was administered to 517 students. Confirmatory factor analysis applied to test the construct validity of the rating scale revealed a good fit of the data with the hypothesized five-factor model. Satisfactory coefficient *H* values for all factors suggested acceptable factor reliability. Cross-validation of the rating scale with two samples showed there was no significant difference in terms of factor loadings and structure. Overall, the psychometric characteristics of the problem quality rating scale turned out to be adequate for measuring quality of problems.

***Keywords*:** Characteristics of problems, Problem-based learning, quality of problem, rating scale, reliability, validity

**INTRODUCTION**

Problems seem to play a pivotal role in Problem-based Learning (PBL) (Majoor, Schmidt, Snellen-Balendong, Moust, & Stalenhoef-Halling, 1990; Williams, Iglesias, & Barak, 2008). In fact, studies point out that besides students' prior knowledge and tutors' performance, the quality of problems has the most significant influence on student learning (Gijselaers & Schmidt, 1990; Schmidt & Gijselaers, 1990; Van Berkel & Schmidt, 2000). Despite the significance ascribed to problems in PBL, surprisingly, there is a lack of validated instruments to measure their quality.

Problems are a set of descriptions of situations or phenomena demanding solutions or explanations, and are usually structured in textual format, sometimes with illustrations, pictures, videos, and simulations (Schmidt, 1983). In PBL, problems trigger the learning process. Problems are purported to achieve the objectives of PBL by engaging students in collaborative work and elaboration, thereby rekindling students' prior knowledge and promoting self-directed learning skills, and consequently leading to construction of new knowledge (Barrows & Tamblyn, 1980; Hmelo-Silver, 2004; Norman & Schmidt, 1992).

In general, there are two approaches to measuring the quality of problems. One approach is to evaluate whether students are able to generate the same learning issues as intended by the curriculum. The degree of congruence between the two is considered to be reflective of problem effectiveness (Dolmans, Gijselaers, Schmidt, & Van der Meer, 1993; Mpofu, Das, Murdoch, & Lanphear, 1997). However, this method has its limitations in the sense that it addresses only one aspect of effective problems – that is, the extent to which a problem leads to the formulation of intended learning issues. In addition, the procedure of comparing the student-generated learning issues with the faculty-intended learning issues may be considered as time consuming and tedious. In a study by Dolmans et al. (1993), 24 expert raters in 12 pairs were to compare a total of 51 learning issues (faculty-intended) with those generated by 120 students for 12 problems. Essentially, what this means is that, given an average of 5 learning issues per problem, across 12 problems and for a total of 120 students, each rater would have to make 7200 comparisons. To reduce the number of comparisons to be made, Dolmans et al. (1993) modified the protocol and allotted one group of 12 students (instead of 120) to each pair of raters. Thus, although, this method provides detailed information about the extent to which a problem leads to the learning predicted issues, the practicality of the method to provide regular feedback about the quality of problems may be limited by the availability of time and resources.

An alternative approach is the administration of a self-report rating scale. To evaluate the quality of a course at the general program level, Schmidt, Dolmans, Gijselaers, and Des Marchais, (1995) developed and validated a 58-item rating scale. Of the 58 items, five items measured the overall quality of all problems in the course. Considering that the measurement scope of the instrument was intended to be at the general program level, it may not be adequate to providing detailed feedback about individual problems.

To assess the complexity and structurednes of PBL problems at individual problem level, Jacobs, Dolmans, Wolfhagen, and Scherpbier, (2003) developed a 12-item rating scale based on Jonassen's theory of problem solving (Jonassen, 2000). When the validity of the rating scale was examined by means of confirmatory factor analysis, results suggested an inadequate fit of the data with the hypothesized two-factor model. Instead, an alteration of the model from the two factor structure to a three-factor yielded a better fit. The altered model consisted of the factors: *too simple*, *too difficult*, and *too well-structured*. These factors were derived from the original two factors by splitting *complexity* into too simple or too difficult, and *structuredness* into too well-structured or too ill-structured, subsequently combining too difficult and ill-structured to form the factor too difficult. Overall, the 12 item-rating scale, encompassing the three factors was concluded to be an adequate instrument to measure the two characteristics of complexity and structuredness. Although the final three-factor model fitted the data reasonably well, it deviates significantly from the initially hypothesized two-factor model and raises concerns about the content validity of the rating scale, since it now measures an extra factor that seems to be conceptually different from what was initially intended.

Marin-Campos, Mendoza-Morals, and Navarro-Hernandez (2004) designed an 18-item rating scale to assess the three aspects of a PBL problem: (1) the extent to which the problem was correctly structured, (2) the extent to which the problem allowed students to carry out the expected learning activities, and (3) the extent to which the allocated time and resources were suitable for the students to work on the problem. Theoretical underpinnings of PBL (Dolmans, Snellen-Balendong, Wolfhagen, & Van der Vleuten, 1997; Rangachari, 1998; Schmidt, 1983) served as the basis for the rating scale design. This rating scale was used to gather feedback on 14 different problems from a group of 28 students longitudinally. Compared to the earlier mentioned studies (Schmidt et al., 1995; Jacobs et al., 2003), this rating scale had the capability to yield more detailed feedback on individual problems. In addition, the internal consistency of the three factors seemed to be adequate when examined by means of Cronbach's alpha test. However there are two points to consider. First, despite the reliability and usefulness of this rating scale to provide detailed feedback on individual problems, its validity remains to be tested. As this study involved only 28 students (from a medical course), validation involving a larger sample by means of factor analysis would still be needed. Secondly, the measurement scope of the rating scale could be extended further. For instance, various core learning activities such as identification of key learning issues, the extent to which the problems encouraged group discussion, and interest triggered by the problem were treated as one factor (the extent to which problem allowed the students to carry out the expected learning activities). Differentiating the various learning activities is likely to provide comprehensive information about the influence of the problem on students' learning.

In summary, the two approaches used currently to assess the quality of problems are (1) comparison of the student-generated learning issues with those intended by the curriculum, and (2) administration of a self-report rating scale to measure a selected set of problem characteristics. Both approaches have their advantages, but when it comes to practical considerations, administering a rating scale seems more feasible. Considering that the existing instruments only addressed a limited number of characteristics (i.e., two or three), we were motivated to develop and validate a more comprehensive problem quality rating scale.

To this end, we first developed a 56-item rating scale measuring eleven characteristics of effective problems in PBL. These characteristics were based on Sockalingam and Schmidt's study (2007) on students' perceptions of problems in PBL and theoretical underpinnings of PBL (e.g., Dolmans, Snellen-Balendong, Wolfhagen, & Van der Vleuten, 1997). Pilot testing of the rating scale showed that the data did not adequately fit the hypothesized 11 factor model and guided us in redesigning the rating scale to a shorter form of 32 items. See the methods section for more on the rationale for the modification of rating scale. The resulting 32-item rating scale was intended to measure the following five problem characteristics: (1) the extent to which the problem leads to formulation of intended learning issues, (2) the extent to which the problem is familiar to students, (3) the extent to which the problem is interesting to students, (4) the extent to which the problem promotes collaborative learning, and (5) the extent to which the problem stimulates critical reasoning. The objective of this study, therefore, was to validate and test the reliability of the 32-item rating scale. To this end, the rating scale was administered to 517 first-year students at a polytechnic in Singapore. Subsequently, confirmatory factor analysis and reliability measures were carried out to examine the psychometric characteristics of the rating scale.

## METHOD

### *Participants*

The sample consisted of 517 participants (58% female and 42% male) with an average age of 18.69 (*SD* = 1.70) years. All participants were enrolled in a first-year general curriculum in the academic year 2007/2008 at a polytechnic in Singapore.

### *Educational Context*

The instructional method used in the polytechnic is PBL for all its modules and programs. In this approach, students work in teams of five members, under the guidance of a tutor. Four to five teams make up a class. Unique to this polytechnic's approach to PBL is that students work on one problem each day (Alwis & O'Grady, 2002). In a week, students work on four to five problems from various modules. A typical day starts with the presentation of a problem. Students discuss in their teams what they know, do not know, and what they need to find out. In other words, students activate their prior knowledge, come up with tentative explanations for the problem, and formulate their own learning goals (Barrows, 1988; Hmelo-Silver, 2004; Schmidt, 1993). Subsequently, a period of self-study follows in which students individually and collaboratively try to find information to address the learning issues (Hmelo-Silver, 2004). At the end of the day the teams come together to present, elaborate, and synthesize their findings.

***Instrument***

*Problem Quality Rating scale*. We first designed a 56-item rating scale to assess eleven characteristics of effective problems. This rating scale was based on Sockalingam and Schmidt's study (2007) on characteristics of problems in PBL and theoretical underpinnings (e.g., Dolmans, Snellen-Balendong, Wolfhagen, & Van der Vleuten, 1997). The eleven characteristics are that problems should (1) be of suitable format (such as length of text and use of visuals), (2) be sufficiently clear, (3) lead to the intended learning issues, (4) be familiar to students, (5) be of appropriate difficulty level, (6) be applicable/relevant (for instance, to other modules/ future work), (7) be interesting to the students, (8) promote self-directed learning, (9) stimulate critical reasoning, (10) encourage teamwork, and (11) trigger elaboration. This rating scale was piloted with 185 first-year students. Confirmatory factor analysis showed the data did not adequately fit the hypothesized factor model. This is not uncommon in developing a new rating scale/questionnaire (Byrne, 2001). We then analyzed the covariance matrix for items that did not contribute significantly to the underlying factors, or were highly correlated. Items that shared higher correlation with other factors; that is items which cross-loaded were combined to form a single factor, taking the conceptual validity into consideration. For instance, three of the characteristics, (1) suitable format of problem (such as length of text and use of visuals), (2) the extent to which the problem is clear, and (3) the extent to which the problem leads to formulation of intended learning issues were combined to form a single factor "the extent to which the problem leads to formulation of intended learning issues". Similarly, two other characteristics; (4) the extent to which problem promotes teamwork, and (5) the extent to which problem triggers elaboration were combined to form a single factor of "the extent to which the problem promotes collaborative learning". Next, items that did not contribute significantly to the underlying latent factor were dropped. This led to too few items for three of the characteristics. Given that initially these characteristics were only represented by four items, the three characteristics had to be excluded. The excluded characteristics were (6) the extent to which the problem promoted self-directed learning, (7) difficulty level of the problem, and (8) the extent to which the problem is applicable/useful. The remaining three characteristics of effective problems, (9) the extent to which the problem is familiar to students, (10) the extent to which the problem is interesting to students, and (11) the extent to which the problem stimulates critical reasoning, were considered to be unique and were used as individual factors in the rating scale. This resulted in a 32-item rating scale, measuring five characteristics of the problems. The five factors of the rating scale are (1) the extent to which the problem leads to formulation of intended learning issues, (2) the extent to which the problem is familiar to students, (3) the extent to which the problem is interesting to students, (4) the extent to which the problem promotes collaborative learning, and (5) the extent to which the problem stimulates critical reasoning. For details of the items, see Appendix A. All items were assessed on a 5-point Likert scale: 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree).

***Procedure***

The rating scale was administered electronically and participants were informed to think about the problem that they had worked on for the day (problem P11) when responding to the rating scale. Participants had fifteen minutes to complete the rating scale.

***Analysis***

First, the 32-items of the rating scale were parcelled, that is combined in groups of two or three based on semantic overlap (Bandalos & Finney, 2001; Little, Cunningham, Shahar, & Widaman, 2002). A total of 14 parcels were formed. Parcelling is a common measurement practice used in latent variable analysis. A parcel can be defined as the average of the two or three indicator items (Little et al., 2002). A detailed description of each of the 14 parcels, accompanied with the indicator items, is given in Appendix C. Next, descriptive statistics for all items and parcels, and correlation matrix for the five factors were generated. Subsequently, confirmatory factor analysis was carried out using AMOS 5 (Arbuckle, 2003) to examine whether the data fitted the hypothesized five-factor model. The analysis was carried out with three different types of samples: First, with an *exploration sample* ($N$ = 209) to conduct an initial analysis of the hypothesized model, and then with a second *construct validation sample* ($N$ = 208) to retest the model and cross-validate the second sample with the first. The cross-validation was done by means of a difference in Chi-square test (Byrne, 2001). As such, the models for the two samples were tested with both unconstrained and constrained factor loadings. Significant differences in Chi-square value between the constrained and unconstrained models in relation to the difference in degrees of freedom reveals the extent to which they differ. After the cross-validation was completed, we retested the five-factor model with the third *main sample,* which is the combined sample of the first two. For all three samples, parameter estimates were generated using maximum likelihood and tests of goodness of fit. Chi-square accompanied by degrees of freedom, sample size, $p$-value, root mean square error of approximation (RMSEA), and comparative fitness index (CFI) were used as indices of absolute fit between the models and the data. The Chi-square is a statistical measure to test the closeness of fit between the observed and the predicted covariance matrix. A small Chi-square value, relative to the degrees of freedom, indicates a good fit (Byrne, 2001). A Chi-square/*df* ratio of less than 3 is considered to be indicative of a good fit (Byrne, 2001). RMSEA is sensitive to model specification and is minimally influenced by sample size and not overly affected by estimation method (Fan, Thompson, & Wang, 1999). The lower the RMSEA value, the better the fit. A commonly reported cut-off value is .06 (Hu & Bentler, 1999). In addition to these absolute fit indices, the comparative fit index (CFI) was calculated. The CFI value ranges from zero to one and a value greater than .95 is conventionally considered a good model fit (Hu & Bentler, 1999).

Finally, Hancock's coefficient *H* was calculated for each of the five factors using the *main sample*. The coefficient *H* is a construct reliability measure for latent variable systems that represents an adequate alternative to the conventional Cronbach's alpha. According to (Hancock & Mueller, 2001) the usefulness of Cronbach's alpha and related reliability measures is limited to assessing composite scales formed from a construct's indicators, rather than assessing the reliability of the latent construct itself as reflected by its indicators. The coefficient *H* is the squared correlation between a latent construct and the optimum linear composite formed by its indicators. Unlike other reliability measures the coefficient *H* is never less than the best indicator's reliability. In other words, a factor inferred from multiple indicator variables should never be less reliable than the best single indicator alone. Hancock recommended a cut-off value for the coefficient *H* of .70.

**RESULTS**

Descriptive statistics were calculated for the items and parcels; no outliers or other abnormalities were observed. The correlations between the five factors ranged from .29 and .65. The mean values and standard deviations, as well as correlation coefficients are presented in Table 1.

**Table 1**

*Descriptive Statistics, Inter-correlations of the Five Factors, and Coefficient H of Main Sample (N = 517)*

| Factor | 1 | 2 | 3 | 4 | 5 | Mean | *SD* | Coefficient *H* |
|---|---|---|---|---|---|---|---|---|
| 1. Learning issue | – | | | | | 3.24 | .60 | .75 |
| 2. Familiarity | .65** | – | | | | 2.99 | .60 | .77 |
| 3. Interestingness | .60** | .56** | – | | | 3.26 | .66 | .77 |
| 4. Collaborative learning | .47** | .29** | .39** | – | | 3.66 | .61 | .78 |
| 5. Critical reasoning | .49** | .38** | .56** | .51** | – | 3.70 | .51 | .66 |

** Correlation is significant at the .01 level.

As a next step we tested whether the factor structure of the five-factor model fitted the data well for all three samples, first, with the exploration sample, followed by the validation sample and Finally with the main sample. The model fit statistics for all three samples are summarized in Table 2.

**Table 2**

*Summary of Goodness-of-Fit Statistics for the Three Samples*

| Sample | *N* | $X^2$ | *Df* | $X^2_{\text{diff}}$ | CFI | RMSEA |
|---|---|---|---|---|---|---|
| Exploration sample | 209 | 76.34 | 64 | 1.19 | .99 | .03 |
| Construct validation sample | 208 | 130.95 | 64 | 2.05 | .94 | .06 |
| Main Sample | 517 | 131.69 | 64 | 2.06 | .98 | .05 |

Note. CFI = comparative fit index; RMSEA = Root mean square error of approximation.

**Table 3**
*Difference in Chi-square Test between the Exploration Sample and the Construct Validation Sample*

| Model description | $X^2$ | *Df* | $X^2_{\text{diff}}$ | $\text{df}_{\text{diff}}$ | Statistical significance |
|---|---|---|---|---|---|
| Hypothesized five-factor model | 207.29 | 128 | _ | _ | _ |
| Model with measurement weights constrained | 214.39 | 137 | 7.11 | 9 | NS** |

**Not significant at the .05 level.

The results demonstrate that the data fitted five-factor model well. The Chi-square/*df* ratio for the main sample, ($N = 517$), was 2.06, $p < .01$, RMSEA = .05 and CFI = .98. All factor loadings, ranging from .59 to .81, were statistically significant and thus contributed significantly to the respective latent variable. The test for invariant factorial structures revealed that there was no significant difference in the underlying factor structure between the exploration sample and the validation sample. See Table 3 for details.

Finally, the reliability of the five factors was assessed using the coefficient *H*, which represents the degree of replicability of a construct based on its measured indicator variables. See Table 1 (eighth column) for details. Values ranged from .66 (Critical reasoning) to .78 (Collaborative learning), with an average of .75. The values are indicative of a moderate to good reliability of the rating scale.

## DISCUSSION

The objective of the present study was to validate and test the reliability a rating scale to measure the quality of individual problems in PBL. To that end, a 32-item rating scale, based on the students' conceptions about five characteristics of effective problems (Sockalingam and Schmidt, 2007) and theoretical underpinnings (e.g., Dolmans, et al., 1997) was developed. The rating scale was tested with 517 first-year students in Singapore context. The factor structure of the rating scale was analyzed by means of confirmatory factor analysis using AMOS 5 (Arbuckle, 2003). Results of the confirmatory factor analysis revealed a good fit of the data with the hypothesized five-factor model. The standardized regression weights of all fourteen parcels were statistically significant, suggesting that the parcels contribute significantly to the underlying latent constructs. The coefficient *H* values for the five factors were satisfactory and indicative of a reasonable reliability. Cross-validation of the rating scale using two samples showed that there was no significant difference in the factor loadings and hypothesized five-factor model between the two groups. In summary, the psychometric characteristics of the 32-item rating scale seemed to be adequate for measuring students' conceptions about the five characteristics of effective problems.

The five factors of the rating scale are (1) the extent to which the problem leads to formulation of intended learning issues, (2) the extent to which the problem is familiar to students, (3) the extent to which the problem is interesting to students, (4) the extent to which the problem promotes collaborative learning, and (5) the extent to which the problem stimulates critical reasoning.

The first factor the extent to which the problem leads to formulation of intended learning issues measures whether the problem instruction is clear, whether the keywords and clues that are embedded in the problem text allow students to identify the intended learning issues, and come up with a logical approach to address the problem. This factor, to some extent, represents Jacob et al.'s (2003) complexity, Marin-Compas et al.'s (2004) two factors on problem structure and problem allowing expected learning activities, and addresses largely the objective of Dolmans' approach to evaluating the effectiveness of problems by means of comparing student-generated learning issues with intended learning issues (Dolmans et al., 1993). Of course, the use of self-report measures has its limitations. The indicator items and parcels used in the rating scale may not be as exhaustive as phenomenological approach. However, considering practicality issues, administering a rating scale is far less time-consuming and feasible.

The second factor, the extent to which the problem is familiar to students, refers to students' familiarity with the context and content of the problem. The familiarity with the problem is the result of past experiences, subject-domain knowledge, and general knowledge. Inclusion of this factor in the rating scale seems reasonable considering the large body of research that suggests that prior knowledge strongly influences learning (Anderson, 1990; Dolmans, Wolfhagen, & Schmidt, 1996; Mamede, Schmidt, & Norman, 2006; Norman & Schmidt, 1992; Schmidt & Boshuizen, 1990; Soppe, Schmidt, & Bruysten, 2005).

The third factor, the extent to which the problem is interesting to students, and the fourth factor, the extent to which the problem promotes collaborative learning, represent the same two factors as in Schmidt's general model of PBL (Schmidt & Gijselaers, 1990). In our case, however, we are more concerned about measuring the student interest and collaborative learning at the problem level to provide detailed feedback on individual problems. As such, the grain-size of our instrument is larger in order to detect differences between individual problems. Interest generated by the problem refers to the level of curiosity and engagement invoked by the problem. Collaborative learning promoted by the problem refers to the extent to which the problem triggers teamwork and elaborations such as brainstorming and discussions. This is also referred to as group functioning in PBL literature.

The fifth and final factor, the extent to which the problem stimulates critical reasoning, refers to the extent to which the problem triggers questioning, stimulates thinking and reasoning, as well as whether the problem allows for multiple solutions. The latter was referred to as structuredness by Jacobs et al. (2003). In our case, however, the fifth factor is broader, and includes questioning, thinking, and reasoning in the context of PBL problems (Kamin, O'Sullivan, Younger, & Deterding, 2001; Tiwari, Lai, So, & Yuen, 2006).

In conclusion, the five factors described above extend the measurement scope of the existing instruments. Besides the characteristics measured by the existing instruments (Schmidt et al., 1995; Jacobs et al., 2003; Marin-Campos, 2004), the problem quality rating scale discussed in this study includes four additional factors (the extent to which the problem is familiar to students, the extent to which the problem is interesting to students, the extent to which the problem promotes collaborative learning, and the extent to which the problem stimulates critical reasoning). This study, therefore, may provide an instrument to measure the quality of problems in a more comprehensive manner than those available at present. Further studies are however needed to establish the instrument’s predictive validity. This demands that the rating scale is administered for number of different problems from different subject-domains and correlated with a corresponding test that adequately determines the predictive validity.

# CHAPTER 5

# STABILITY AND INTER-RATER RELIABILITY IN STUDENT RATINGS OF PROBLEM QUALITY IN PBL

## ABSTRACT

This study investigated the inter-rater reliability and the stability of ratings over time, and across different groups of students in using a problem quality rating scale. To this end, 244 students were asked to rate a problem that they had worked on using the rating scale. Two problems were used in the study. The intra-class coefficient for the two problems averaged .87 and .92, suggesting that students were able to judge the quality of the problems in terms of their most important characteristics. In addition, temporal stability of the students' ratings was studied at various points in time. The results showed variance of ratings over time, demonstrating that student judgment of the quality of problems is dependent on the amount of experience that students have with a particular problem. As for the stability of ratings across different ability groups of students, that is low, medium, and high scorers on (1) a knowledge tests from a previous module, (2) a knowledge tests from the module under study, and (3) tutors' observation of students' actual learning activities in the module under study, the results generally showed invariance in ratings on four of the five characteristics. The exception was that groups identified based on (1) scores in the previous module, and (2) tutors' observation of students' learning activities varied in their ratings on the extent of problem interestingness. Overall, the results demonstrate that the rating scale enables students to rate the quality of the problems presented in an accurate fashion. In addition, scores do not appear to be influenced

by the students' abilities although the phase of the learning cycle in which problem quality is measured seems to make a difference.

***Keywords*****:** Characteristics of problems, inter-rater reliability, problem-based learning, quality of problem, rating scale, stability

**INTRODUCTION**

The quality of problems in Problem-based Learning (PBL) has been suggested to be a crucial factor for students' learning (Gijselaers & Schmidt, 1990; Schmidt & Gijselaers, 1990; Schmidt & Moust, 2000; Van Berkel & Dolmans, 2006; Van Berkel & Schmidt; 2000). This raises the question as to whether the ratings of the quality of problems are stable regardless of who is rating the problem and when it is being rated. If the ratings of the quality of problems are unstable, then the assessment of problem quality can be considered to be unreliable.

PBL, as the name implies, is based on problems, and is founded on the tenet that learning is student-centered, collaborative and engaging (Hmelo-Silver, 2004; Norman & Schmidt, 1992). The problem in PBL refers to the instructional material given to students to trigger their learning process, and is usually a set of description of phenomena or situations which require explanations or resolution. When presented with the problem, students work in their tutorial groups to examine and discuss it. In this process, they identify what they know and do not know about the problem, and determine the issues that need further exploration. Unresolved questions and doubts generated from the discussion serve as guidelines for further exploration in the subsequent period of self-study. After the period of self-study, they reconvene with their team to relate their findings, and synthesize their new understanding about the problem (Hmelo-Silver, 2004). Overall, learning in PBL is largely student-owned. Rather than providing knowledge directly, the tutor in PBL facilitates the learning process by monitoring the students' learning activities, diagnosing any issues or challenges faced by students, and providing necessary feedback at appropriate instances (Neville, 1999).

Studies comparing the effect of the quality of problems and role of tutors on students' learning suggest that the former is more influential (Gijselaers & Schmidt, 1990; Schmidt & Gijselaers, 1990; Schmidt & Moust, 2000; Van Berkel & Dolmans, 2006; Van Berkel & Schmidt; 2000). These studies show that the quality of problems has a positive influence on the extent to which students engage in collaborative learning and the extent to which they spend time in self-study. These learning activities in turn positively influence their academic achievements and interest in the subject matter. The implication of this is that a good problem is likely to lead to better learning. Taking a different perspective and assuming that typically a single problem is facilitated by a number of tutors across various small-group tutorials, one can argue that the number of students per problem is higher than the number of students per tutor. Hence, it is reasonable to infer that a problem is likely to have a wider influence than a tutor, and therefore its quality is even more important than the individual tutor's quality. In sum, the quality of problems seems to play a significant role in students' learning. Despite its suggested importance, there are only a limited number of studies which explore ways to assess problem quality (Jonassen & Hung, 2008).

Generally, two approaches are employed to measure the quality of problems. Both these approaches are based on the principle that an effective problem should meet the expected targets or objectives it was intended for. One approach is to compare the student-generated learning issues with the faculty-intended learning issues (Dolmans, Gijselaers, Schmidt, & Van der Meer, 1993; Mpofu, Das, Murdoch, & Lanphear, 1997). The advantage of this approach is that the measurement of the quality of problem is independent of the learners. On the other hand, only one aspect of the problem quality; the extent to which the problem leads to the intended learning issues, is assessed. Furthermore, this approach can be time-consuming and resource-intensive.

The second approach addresses the limitations mentioned earlier. In this approach, students' perceptions about the characteristics of problems are collated by means of a rating scale. For instance, Jacobs, Dolmans, Wolfhagen, and Scherpbier (2003) developed, and assessed the validity of a 12-item rating scale measuring two characteristics of problems; the extent of problem complexity and "structuredness". Complexity refers to the extent of cognitive actions and steps needed to answer the problem while structuredness refers to the multiplicity in solutions. They found that students could differentiate the extent of problem simplicity and well-structuredness, but not the extent of problem complexity and ill-structuredness. Hence the authors merged the later two sub-factors into one factor known as problem difficulty. They concluded that the rating scale can be used to detect whether PBL problems are too simple, too well-structured, or too difficult in students' opinions.

Marin-Campos, Mendoza-Morals, and Navarro-Hernandez (2004) developed an 18-item rating scale with the objective to measure three characteristics of PBL problems. The three problem characteristics assessed were; (1) the extent to which the problem leads to learning activities such as self-directed study, and collaborative learning, (2) how the problem is structured/represented to lead the students to the intended content, and (3) the extent to which time and resources were needed to tackle the problem. They administered the rating scale to students to gather feedback on 14 problems and found that measurement using the rating scale was reliable.

Soppe, Schmidt, and Bruysten (2005) and Verkoeijen, Rikers, Te Winkel, and Van der Hurk (2006) also used the rating scale approach to assess specific characteristics of problems. In addition, they investigated the influence of these characteristics on students' learning. Soppe et al., (2005) used a 12-item rating scale to assess students' perceptions about the level of problem familiarity and interestingness, and compared that with (1) quality and quantity of learning issues generated by students, (2) students' achievement, and (3) time taken for self-study. Two groups of students, one working on the "familiar" version and the other working on the "unfamiliar" version of the same problem, were administered the rating scale. The results suggested that the students working on the problem with the familiar context perceived it to be more familiar than the problem with not-so-familiar context. In addition, they reported that the familiar problem was interesting. However, no significant difference was found between the two groups of students in terms of the learning issues generated, achievement, and the amount of time taken for self-study.

Verkoeijen, et al. (2006) investigated the extent to which goal specification in a problem influenced students' learning by administering either a "goal-free" or a goal-specified" problem to two groups of students. The goal-free problem did not specify any goal whereas the goal-specified problem stated the goal needed to be achieved by students. The authors postulated that the students working on goal-free problem would read more articles, and spend longer time in self-study. In addition, they expected that these students would spend more time in reporting phase. To measure the learning outcomes, they administered a short rating scale which assessed the quality of the PBL cycle in students' perceptions. Although the authors did not report the exact items in the rating scale, they explained that the items measured (1) the depth and quality of discussion and reporting phases, (2) the extent of elaboration during the discussion phase, and (3) the students' perceived mastery in the subject matter. These measures were then compared with tutors' record of time taken by students for discussion and reporting phase, and students' record of time spent on self-study and the number of articles studied. The results showed that the students working on the goal-free problem read more articles, spent more time in studying as well as reporting the findings. Furthermore, the students working on the goal-free problem perceived to have a higher level mastery of the subject matter compared with the students working on the goal-specified problem.

Overall, the advantages of the rating scale approach are that first, it allows the measurement of characteristics other than the extent to which the problem leads to the intended learning issues (e.g., the extent to which the problem is interesting). Second, administration of the rating scale can be considered to be less time-consuming and resource-intensive. However, the earlier mentioned rating scales are generally not validated, with the exception of the rating scale by Jacobs et al., 2003. This rating scale, however, confined itself to two characteristics of problems; the extent of problem complexity and structuredness.

Based on the literature in PBL and results from focus group studies with students and tutors on effective problems (Sockalingam & Schmidt, 2008), Sockalingam, Rotgans, and Schmidt (2009) devised a 32-item problem quality rating scale measuring the following five problem characteristics: the extent to which the problem leads to the intended learning issues, the extent to which the problem is familiar to students, the extent to which the problem is interesting to students, the extent to which the problem promotes collaborative learning, and the extent to which the problem stimulates critical reasoning. To assess the reliability and validity of the rating scale, they administered the rating scale to 517 first-year students in a problem-based curriculum. The adequacy of the five measures was determined using confirmatory factor analysis. The results suggested a fairly good fit of the hypothesized model to data. The coefficient-*H* values of the five measures ranged from .66 to .78 indicating good internal consistency of the five measures. In sum, the rating scale measuring the five characteristics of problem quality was shown to be valid and reliable in terms of internal consistency.

Although Sockalingam et al. (2009) had shown that the measures of the five problem characteristics of problems are reliable in terms of internal consistency, their results did not suggest whether these different raters would produce similar ratings of the different problem characteristics (inter-rater reliability) and whether these measures were stable over time and across different student groups. According to Cohen, Manion, and Morrison (2000) the three dimensions of reliability are internal consistency, stability, and equivalence. Internal consistency measure of the rating scale indicates the extent of congruency and consistency between the different items of a single scale. On the other hand, stability refers to the measure of consistency over time and across different raters. It is possible that the rating scale is reliable in terms of internal consistency, but only when measured at a certain time-point or when administered to a specific group of students. For instance, high achievers may evaluate the problems differently from low achievers.

In testing the temporal stability in students' ratings over the period of the PBL learning cycle, one hypothesis is that the ratings are not influenced by the time-point of rating scale administration as students are responding on the same problem. If this is true, it means that students do not let their judgment of the quality of problem be influenced by how much experience they have with the problem. However, it is possible that students' experience with the problem has an impact on their ratings of the quality of the problem. For instance, students who have just read the problem may not find the problem to be as interesting as when compared with the interest level after discussing with teammates. Several studies show that group functioning has a positive and significant influence on level of interest (Dolmans & Schmidt, 2006; Gijselaers & Schmidt, 1990; Schmidt & Gijselaers, 1990; Schmidt & Moust, 2000). Hence students' perceptions of the extent to which the problem is interesting may be higher after the initial discussion than after reading the problem. It is also possible that students may report lower interest after the final discussion if their epistemic curiosity is satisfied (Schmidt, 1993). Likewise variations may also be observed in students' ratings of the other problem characteristics.

In addition, we wanted to know if students' achievements in knowledge tests and the quality of their learning activities while working on the problem influenced how they rated the problem. It is possible that students who scored high in a previous module feel more positive/confident, and rate the subsequent problem to be good. It could also be that students who had not received good scores rate the subsequent problem to be bad. Studies have shown that students' grades influence their evaluation of course work and teaching (Brown, 1976; El Ansari & Oskrochi, 2004; Engdahl, Keating, & Perrachione 1993; Feldman, 1976). On the other hand, even if students do not know their grades, it is possible that high achievers and low achievers rate the problems differently. Boud and Falchikov (1989) reported that high achievers are more competent in self-assessment compared with low achievers. This motivated us to find out if students' achievement in knowledge tests reflected how they rated a given problem.

In PBL, knowledge tests can be considered as a summative assessment of students' learning and tutors' observation of their learning activities can be considered as formative assessment of their learning. As the knowledge tests may not be absolutely representative of what happens during the learning process, we wanted to investigate whether students exhibiting different level of learning activities perceived the problem quality differently. For instance, it may be possible that students who demonstrate high level of collaborative work rate the problem to be higher on the problem characteristics of "the extent to which the problem promoted collaborative learning". Dolmans and Wolfhagen (2005) reported that students' perceptions of their group productivity were highly correlated with their perceptions of the effectiveness of their learning unit. However, in their study, effectiveness of learning unit was considered at the general level and not on specific characteristics of the problem, and the measurement of students' learning activities were students' perceptions rather than tutor's observations. Hence, we wanted to know if students observed to have demonstrated different levels of learning activities varied in their ratings of specific problem characteristics.

In sum, the objective of this study was to assess the inter-rater reliability and the stability of the ratings over time, and across different groups of students in using the problem quality rating scale. More specifically, this research set out to investigate (1) the intra-class correlation of the students' ratings of two problems, (2) the temporal stability of the ratings across various time points during the learning cycle, and (3) the stability of ratings across different ability groups of students. To this end, the rating scale was administered to 244 first-year students from a psychology module for two problems and the students were asked to respond on the problem they worked on, at different points in time.

## METHOD

### *Participants*

The data were collected in the academic year 2007-2008. All first-year students were involved in the study ($N$ = 244). Response rate was 91%. Mean age of the participants was 19 years.

### *Educational context*

The research was conducted in the Personality Psychology module of the psychology curriculum at Erasmus University in the Netherlands. PBL is the dominant educational method used in this institution. PBL process adopted here is as described earlier in the introduction section; the learning process encompasses two meetings with a self-study period of two days in-between meetings. The students work in groups of 8-10 on the same problem through out the two meetings and self-study time.

***Measures***

*Problem Quality Rating Scale*. A problem quality rating scale developed and validated by Sockalingam et al., (2009) was used to assess the quality of problems. The 32-item problem quality rating scale measures the following five problem characteristics: (1) the extent to which the problem leads to the intended learning issues, (2) the extent to which the problem is familiar to students, (3) the extent to which the problem is interesting to students, (4) the extent to which the problem promotes collaborative learning, and (5) the extent to which the problem stimulates critical reasoning. The coefficient-*H* values of the five measures were shown to be reliable. According to Hancock and Mueller (2001), the usefulness of Cronbach's alpha and related reliability measures is limited to assessing composite scales formed from a construct's indicators, rather than assessing the reliability of the latent construct itself as reflected by its indicators. Unlike other reliability measures, the coefficient *H* is never less than the best indicator's reliability. The details of items in the problem quality rating scale can be found in Appendix C. Students were asked to rate each of these items on a 5-point Likert scale of 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree). The rating scale instruction stated that there were no right or wrong answers to the items. No information was given about the constructs underlying the rating scale.

***Procedure***

Students rated two problems from the Personality Psychology module using the rating scale. The two problems were (1) Genetic and biochemical roots of personality and (2) (In) stability of human behaviour. For the first problem, the rating scale was administered at one of the three time-points; (1) after reading the problem, (2) after the initial discussion, (3) and after the final discussion. Students who were asked to respond after reading the problem were presented with a shorter version of the rating scale. As we felt that certain items in the rating scale such as “The problem captivated my attention throughout the day”, “Our teamworked efficiently” can only be answered after working on the problem, we excluded these items from the rating scale, resulting in a shorter version of 16-items. The shorter rating scale measured three of the five problem characteristics; the extent to which the problem leads to the learning issues, the extent to which the problem is familiar, and the extent to which the problem is interesting. The items used in the shorter version are indicated in Appendix C. As for the second problem, the rating scale was administered at one of the two time-points; (1) after reading the problem, and (2) after the initial discussion. As before, students responding after reading the problem used the shorter version and students responding after the initial discussion used the longer version of the rating scale. Students took 10-15 minutes to respond on the rating scale. They were not informed how the data were going to be analyzed. Average measures of the problem characteristics were computed at the individual student level.

*Categorization of students based on knowledge.* As a routine practice in the curriculum, students are required to attempt a knowledge tests at the end of every module. The knowledge tests consists of multiple-choice questions and open-ended questions that assess students’ understanding of the concepts pertaining to the respective module. Students receive an individual score, ranging from 0 to 10. Students' knowledge tests scores on a previous module called "Social Psychology", and module under study, "Personality Psychology" were used in the data analysis. At the point of study, students were aware of their knowledge tests score for the previous module but not the module under study. These students were subsequently categorized into "low", "medium" and, "high" groups based on their scores on each of the two knowledge tests. Students who scored below the sum of mean and half the standard deviation were grouped into the "low" category. Students who scored above the sum of mean and half the standard deviation for that particular variable were grouped into the "high" category. Students who scored in the range between "low" and "high" were categorized into the "medium" group.

*Categorization of students based on their actual learning activities.* As part of the regular assessment, tutors observe and rate the students on a number of criteria using a tutor rating scale measuring students' learning activities. The criteria includes (1) how well the students prepared themselves with respect to the subject matter studied, (2) how active, motivated and participative the students were in the group activities, and (3) how well the students fulfilled their roles as chair and scribe. Ratings on these dimensions range from 1 (student did not show these activities at all) to 5 (student showed these activities to a large extent). Average score of the three criteria is awarded to students as the tutor's score for students' learning activities. Loyens, Rikers, and Schmidt (2007) showed that measurement of the students' learning activities using the tutor rating scale is highly reliable. As before, the students were categorized into "low", "medium", and "high" groups based on the mean and standard deviation.

***Analysis***

First, intra-class correlations were computed across the different raters for each of the five problem characteristics for the two problems. To determine the temporal stability of the students' ratings, a series of multivariate analysis of variance were carried out. The independent variable used was the time-point of rating scale administration while the dependent variables used were the average ratings on the various problem characteristics. For the first problem on "Genetic and biochemical roots of personality", the students' ratings measured at two time points of the learning cycle; after the initial discussion and after the final discussion were compared for all five characteristics. In addition, we compared the ratings for the first problem across three time points (after reading the problem, after the initial discussion, and after the final discussion) for three problem characteristics; the extent to which the problem led to the learning issues, the extent to which the problem was interesting, and the extent to which the problem was familiar. For the second problem on "(In) stability of human behaviour", we compared the ratings at two time points (after reading the problem and after the initial discussion) for the three characteristics. To examine the stability of ratings across different ability groups of students, we again utilized multivariate analysis of variance tests. This time, the independent variables were (1) grouping of students based on the knowledge tests score from a previous module (three groups) , (2) grouping of students based on the knowledge tests score from the module under study (three groups), and (3) grouping of students based on tutors' observation of students' learning activities (three groups). The dependent variables used were the five problem characteristics.

**RESULTS**

Preliminary assumption testing was first carried out to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices and multicollinearity. No serious violations were noted. To determine whether students varied in their ratings of the five problem characteristics, intra-class correlation measures were computed for the five characteristics for the two problems. Table 1 shows the intra-class correlation measures. Average of the intra-class correlation for the five characteristics on this problem was found to be .87 for the first problem and .92 for the second problem.

**Table 1**

*Intra-class Correlation Measures of Students' Ratings on the Problem Quality Rating Scale*

| Problem characteristics | Problem 1: Genetic and biochemical roots of personality | Problem 2: (In)stability of human behaviour |
|---|---|---|
| 1. Led to learning issues | .72 | .87 |
| 2. Was familiar | .90 | .97 |
| 3. Was interesting | .95 | .97 |
| 4. Stimulated critical reasoning | .80 | .84 |
| 5. Promoted collaborative learning | .97 | .93 |

To examine the temporal stability of the students' ratings, series of multivariate analysis were carried out. This was done first for the problem on "Genetic and biochemical roots of personality" whereby we compared the ratings across two time-points; after the initial discussion and after the final discussion, for all five characteristics. The results show no significant difference on three of the characteristics; the extent to which the problem was interesting, familiar, and promotes collaborative learning. However, significant differences were found in the extent to which the problem led to the learning issues [$F(1, 139) = 18.08$, $p = .00$, partial eta squared = .12] and the extent to which the problem stimulated critical reasoning [$F(1, 139) = 7.43$, $p = .01$, partial eta squared = .05] (See Table 2).

Next, we compared the students' ratings for the same problem across three time points (after reading the problem, after the initial discussion, and after the final discussion) for three of the five characteristics; the extent to which the problem led to the learning issues, the extent to which the problem was interesting, and the extent to which the problem was familiar. The results showed that there was significant difference in the extent to which the problem led to the learning issues [$F(2, 208) = 10.61$, $p = .00$, partial eta squared = .09]. Post-hoc analysis showed that there ratings measured at the two time points; after reading the problem ($M$ =3.31, $SD$=.59) and after the initial discussion ($M$ =3.11, $SD$=.32) formed one sub-group while ratings measured after the final discussion formed another sub-group ($M$ =3.60, $SD$=.59). The general trend for all three characteristics was that there was a dip in the ratings in the first half of the PBL cycle (from the time- point of "after reading problem "to "after the initial discussion") and an increase in the ratings in the second half of the PBL cycle (from the time-point of "after the initial discussion" to "after the final discussion").

In addition, we compared the students' ratings for the second problem on "(In) stability of human behaviour" across two time points; after reading the problem and after the initial discussion, for the three characteristics as before. There were significant differences in the extent to which the problem was interesting [$F(1, 220) = 5.13$, $p = .02$, partial eta squared = .02], and the extent to which the problem was familiar to students [$F(1, 220) = 4.18$, $p = .04$, partial eta squared = .02]. No significant difference was found on the extent to which the problem led to the learning issues. Ratings on the extent to which problem was interesting and familiar were found to be lower after the initial discussion than after reading the problem (see Table 2).

To investigate the stability of ratings across different ability groups of students, one-way MANOVA tests were carried out with the students' ratings on the second problem "(In) stability of human behaviour". The different groups of students were "low", "medium", and "high" scorers on knowledge tests scores from (1) a previous module, (2) module under study, and also (3) tutors' observation of students' learning activities. Table 3 and 4 present the descriptive statistics and MANOVA results.

No significant difference was found in the ratings on the five problem characteristics between the three levels of achievers from the module under study (Personality Psychology). Similarly, no significant difference was found in the ratings between the three levels of achievers from the previous module and between groups of students demonstrating different levels of learning activities on four of the five characteristics. The exception was that these groups varied in their ratings on the extent of problem interestingness. Students who scored high in the previous module rated the problem to be more interesting ($M$ = 3.28, $SD$ = .52) than students who scored lower ($M$ = 2.88, $SD$ = .58). Post-hoc analysis showed that low scorers in the previous module differed from the medium and high scorers. Also, students who demonstrated high level of learning activities rated the problem to be more interesting ($M$ = 3.32, $SD$ = .61) than the medium ($M$ = 2.89, $SD$ = .59) and low groups ($M$ = 3.09, $SD$ = .61). Post-hoc analysis showed that the "medium" and "high" groups clearly differed. Interestingly, "low" and "high" groups did not differ.

**Table 2**

*Students' Ratings on the Problem Quality Rating Scale at Various Time Points for the Two Problems*

| Problem | Problem characteristics | | Time of rating scale administration | | | | | | *F* | *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | After initial discussion (*n = 53)* | | After final discussion (*n = 87)* | | | |
| | | | | | mean | *SD* | mean | *SD* | | |
| Genetic and biochemical roots of personality | 1. | Led to learning issues | | | 3.12 | .79 | 3.61 | .57 | 18.08 | .00* |
| | 2. | Was familiar | | | 2.77 | .70 | 2.91 | .60 | 1.59 | .21 |
| | 3. | Was interesting | | | 3.31 | .82 | 3.39 | .59 | .38 | .54 |
| | 4. | Stimulated critical reasoning | | | 3.17 | .32 | 3.01 | .35 | 7.43 | .01* |
| | 5. | Promoted collaborative learning | | | 3.19 | .84 | 3.20 | .59 | .01 | .91 |
| | | | After reading (*n = 61 )* | | After initial discussion (*n = 63)* | | After final discussion (*n = 87)* | | . | |
| | | | mean | *SD* | mean | *SD* | mean | *SD* | | |
| Genetic and biochemical roots of personality | 1. | Led to learning issues | 3.31 | .59 | 3.11 | .32 | 3.60 | .59 | 10.61 | .00* |
| | 2. | Was familiar | 2.92 | .58 | 2.81 | .84 | 2.91 | .60 | .65 | .52 |
| | 3. | Was interesting | 3.81 | .68 | 3.52 | .82 | 3.61 | .68 | 2.53 | .08 |
| | | | After reading (*n* = 103*)* | | After initial discussion (*n* = 119*)* | | | | | |
| | | | mean | *SD* | mean | *SD* | | | | |
| (In) stability of human behaviour | 1. | Led to learning issues | 3.11 | .65 | 3.04 | .68 | | | .74 | .39 |
| | 2. | Was familiar | 2.90 | .53 | 2.76 | .47 | | | 4.18 | .04* |
| | 3. | Was interesting | 3.67 | .75 | 3.44 | .72 | | | 5.13 | .02* |

** Significant at .05 level*

**Table 3**

*Ratings on the Problem Quality Rating Scale by Students of Different Achievement Levels in Knowledge Tests*

| Problem characteristics | Achievement on a previous module | | | | | | *F* | *p* | Achievement on the module under study | | | | | | *F* | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Low ($n$ = 31) | | Medium ($n$ = 48) | | High ($n$ = 40) | | | | Low ($n$ = 31) | | Medium ($n$ = 50) | | High ($n$ = 38) | | | |
| | mean | *SD* | mean | *SD* | mean | *SD* | | | mean | *SD* | mean | *SD* | mean | *SD* | | |
| 1. Led to learning issues | 3.02 | .76 | 3.04 | .55 | 3.20 | .60 | .99 | .38 | 2.97 | .61 | 3.16 | .62 | 3.10 | .65 | .87 | .42 |
| 2. Was familiar | 2.71 | .51 | 2.75 | .46 | 2.81 | .47 | .38 | .69 | 2.69 | .47 | 2.82 | .52 | 2.74 | .41 | .80 | .45 |
| 3. Was interesting | 2.88 | .58 | 3.08 | .68 | 3.28 | .52 | 3.89 | .02* | 3.14 | .49 | 3.14 | .66 | 3.00 | .66 | .64 | .53 |
| 4. Stimulated critical reasoning | 3.18 | .43 | 3.17 | .49 | 3.17 | .39 | .01 | .99 | 3.15 | .46 | 3.15 | .45 | 3.22 | .43 | .28 | .76 |
| 5. Promoted collaborative learning | 3.19 | .59 | 3.33 | .59 | 3.39 | .48 | 1.22 | .30 | 3.44 | .51 | 3.30 | .51 | 3.23 | .65 | 1.22 | .30 |

* *Significant at .05 level*

**Table 4**

*Ratings on the Problem Quality Rating Scale by Students Observed to have Demonstrated Different Levels of Learning Activities*

| Problem characteristics | Different level of students' learning activities | | | | | | $F$ | $p$ |
|---|---|---|---|---|---|---|---|---|
| | Low ($n$ = 45) | | Medium ($n$ = 38) | | High ($n$ = 36) | | | |
| | mean | *SD* | mean | *SD* | Mean | *SD* | | |
| Led to learning Issues | 3.15 | .66 | 2.90 | .60 | 3.22 | .57 | 2.89 | .06 |
| Was familiar | 2.78 | .51 | 2.66 | .46 | 2.84 | .44 | 1.40 | .25 |
| Was interesting | 3.09 | .61 | 2.89 | .59 | 3.32 | .61 | 4.57 | .01* |
| Stimulated critical reasoning | 3.09 | .47 | 3.24 | .45 | 3.20 | .39 | 1.28 | .28 |
| Promoted collaborative learning | 3.33 | .58 | 3.21 | .58 | 3.38 | .48 | .86 | .43 |

* *Significant at .05 level*

**DISCUSSION**

This study aimed to investigate inter-rater reliability, temporal stability, and stability in view of differences in student ability of ratings of problem quality using a problem quality rating scale (Sockalingam et al., 2009). The rating scale measures five characteristics of problems: the extent to which the problem leads to the intended learning issues, the extent to which the problem is familiar to students, the extent to which the problem is interesting to students, the extent to which the problem promotes collaborative learning, and the extent to which the problem stimulates critical reasoning. To this end, 244 students were asked to rate a problem that they had worked on using the rating scale. Two problems were used in the study. The intra-class coefficient for the two problems averaged .87 and .92, suggesting that students were able to judge the quality of the problems in terms of their most important characteristics. In addition, temporal stability of the students' ratings was studied at various points in time. The results showed variance of ratings over time, demonstrating that student judgment of the quality of problems is dependent of the amount of experience that students have with a particular problem. As for the stability of ratings across different ability groups of students, that is low, medium, and high scorers on (1) a knowledge tests from a previous module, (2) a knowledge tests from the module under study, and (3) tutors' observation of students' actual learning activities in the module under study, the results generally showed invariance in ratings on four of the five characteristics. The exception was that groups identified based on (1) scores in the previous module, and (2) tutors' observation of students' learning activities varied in their ratings on the extent of problem interestingness. Overall, the results demonstrate that the instrument developed enables students to rate the quality of the problems presented in an accurate fashion. In addition, scores do not seem to be influenced by the abilities of the students working on these problems although phase of the learning cycle in which problem quality is measured seems to make a difference.

The average intra-class correlation measure for the first problem on “Genetic and biochemical roots of personality” was found to be .87, while that for the second problem on “(In)stability of human behaviour”was found to be .92, suggesting high level of agreement between the raters on the problem quality. This measure, however, indicates the overall agreement between the raters on the five problem characteristics. Differences in ratings due to time-point of rating scale administration and between groups of students who differ in their learning activities and knowledge tests may not be revealed by this measure.

Indeed, the time point of administration seems to have an effect on the ratings on problem characteristics. Students rated the first problem to have led to more learning issues after the final discussion ($M$ = 3.61, $SD$ = .57) than after the initial discussion ($M$ = 3.12, $SD$ = .79). On the other hand, they felt that the problem had resulted in lower level of critical reasoning after the final discussion ($M$ = 3.01, $SD$ = .35) than after the initial discussion ($M$ = 3.17, $SD$ = .32).This could be possibly because students’ perceptions about their learning changes with time. For instance, students may brainstorm and come up with several unanswered queries during the initial discussion. Typically, these unresolved queries help to direct the students in their self-directed learning activities (Hmelo-Silver, 2004). Hence, after the initial discussion they may feel that the problem has not yet led to sufficient learning issues. However, by the end of the final discussion, students may feel that the problem has led to satisfactory level of learning issues as most of their queries are answered through self-directed and collaborative learning (Hmelo-Silver, 2004). This may also explain why students perceived the problem to have stimulated more critical reasoning after the initial discussion than after the final discussion. Students' queries may have been answered resulting in satisfaction of their epistemic curiosity by the end of the final discussion (Schmidt, 1993).

This idea is supported by other studies in literature. For instance, a study by Dolmans, Schmidt, and Gijselaers (1995) shows that students identified 64% of the faculty-intended learning issues by the end of the learning session. Another finding was that the learning issues generated during group discussion were not the only guiding force directing the students in their choice of what to study. They found that the students may decide to study on topics not directed by the learning issues. This means that what students had learnt right after reading the problem is likely to be different from what they learn by the end of the final discussion. Nonetheless, students would have learnt more by the final discussion.

Similar results were also found on the shorter rating scale. Interestingly, students' ratings on the extent to which the problem led to the learning issues was found to be higher right after reading the problem (*M* = 3.31, *SD* = .59) than after the initial discussion (*M* = 3.11, *SD* = .32). This could be attributed to students' (mis) conceptions of what they thought the problem was about right after reading the problem. However, the initial discussion could have revealed several areas of knowledge gaps, resulting in the lowering of their ratings on the extent to which the problem led to the learning issues after the initial discussion.

Although one would expect that more unresolved queries and perceived high level of critical reasoning at the end of final discussion would have corresponded with an increase in the interest level, the interest level was found to be unaffected over time for this problem. Students reported similar level of interest, after the initial discussion (*M* = 3.31, *SD* = .82), and after the final discussion (*M* = 3.39, *SD* = .59). Similarly, students' ratings on the extent to which the problem was found to be familiar (after the initial discussion; *M* = 2.77, *SD* = .70, and after the final discussion; *M* = 2.91, *SD* = .60), and to have promoted collaborative learning was also found to be invariant (after the initial discussion; *M* = 3.19, *SD* = .84, and after the final discussion; *M* = 3.20, *SD* = .59).

In contrast, ratings on the second problem showed (1) no significant difference in the ratings on the extent to which the problem led to the learning issues (after reading the problem; *M* = 3.11, *SD* = .65 , and after the initial discussion; *M* = 3.04, *SD* = .68), and (2) significant differences in the extent to which the problem was found to be interesting (after reading the problem; *M* = 3.67, *SD* = .75 , and after the initial discussion; *M* = 3.44, *SD* = .72), and familiar (after reading the problem; *M* = 2.90, *SD* = .53 , and after the initial discussion; *M* = 2.76, *SD* = .47). These differences in the ratings of the two problems over time lead to three inferences. First, the results reveal that time-point of rating scale administration is a potential source of variance in students' ratings of the problem quality. The ratings seem to be the lowest after the initial discussion than after reading of the problem and after the final discussion for both the problems. Nevertheless, generalizability of this observation needs to be further tested. The results could also mean that students' experience with the problem is critical to their assessment of the problem quality. Most studies assessing the quality of problems do so after students have worked on the problem (Jacobs et al., 2003; Marin-Campos et al., 2004; Soppe et al., 2005; Verkoeijen et al., 2006). As far as we know, the present study may be the first exploration of whether the time point of problem quality assessment affects the students' ratings. The finding that students' ratings vary over time during the PBL cycle provides new insights. The results support the notion that engagement in problem is needed for assessment of the problem quality.

Second, these results also support the validity of the rating scale. Studies testing for reliability of ratings aim to show high reliability values. Conversely, consistent high values regardless of context may mean that the measure is invalid. For instance, students are generally expected to have learnt more at the end of the final discussion compared to the initial discussion. If the rating scale to assess problem quality is not able to discern this difference, then there is a possibility that the rating scale is not valid. Students' ratings on the two problems at different time points indicate that the rating scale is able to detect the variations, therefore adding support to its validity.

Third, variation in problems seems to influence students' ratings. This is in line with studies by Marin-Campos et al., 2004; Soppe et al., 2005; Verkoeijen et al., 2006. However, it is not the objective of this research to decipher how exactly the quality of problem influences the students' ratings from this study. This presents itself for further research.

As for the stability of ratings across different ability groups of students, regardless of the scores obtained in module under study, students rated the problem similarly. A possible reason for the invariance of ratings could be that the students in our study are able to assess the problem quality equally well. The results also show that students who scored higher in the previous module rated the problem to be more interesting. This could be because these students were aware of their grades. Literature suggests that the students' rating of curriculum could be influenced by their grades (Brown, 1976; El Ansari & Oskrochi, 2004; Engdahl, Keating, & Perrachione 1993; Feldman, 1976).

Students who were observed to have demonstrated higher learning activities also rated the problem to be more interesting. From literature, we know that task engagement (that is demonstration of learning activities) is known to be correlated with interest in learning activities (Krapp, Hidi, & Renninger, 1992). Interest in learning can be distinguished as due to personal interest and situational interest. Personal interests are specific to individuals and tend to endure and evolve over time and across different situations (Renninger, 2000). Situational interests refer to interests that are evoked by the interestingness of a situation or context (Krapp et al., 1992). In our case, we can define the interest triggered by the problem to be situational interest. As the three groups of students who were identified to be demonstrating different levels of learning activities were presented with the same problem, we can assume that the situational interest is not varied. Hence, it is logical to infer that the students' personal interest in the problem could have led them to varying extent of learning activities. In that case, it seems reasonable that students observed to be highly engaged rated the problem as highly interesting (Krapp et al., 1992). What strikes is that students rated to be "low" in learning activities also rated the problem to be highly interesting. One possibility is that tutors could have rated the students' learning activities with reference to the faculty-intended learning issues. Studies show that often students explore topics that may be related but not necessarily directed at answering the problem (Dolmans et al., 1993; Dolmans et al., 1995; Mpofu et al., 1997). Hence, such students may have been rated "low" by tutors.

In explaining the difference in students' ratings of problem interestingness, it is also possible that students' grades in the previous module boosts their confidence, resulting in a higher perceived ability (Parsons, Croft, & Harrison, 2009). Greene and Miller (1996) show that perceived ability is positively correlated with meaningful cognitive engagement. Cognitive engagement on a given problem in turn could reflect the level of interestingness that the students associate with the problem (Krapp et al., 1992). A correlation of the tutor's score of students' learning activities and their score in the previous module was found to be .14 (correlation is significant at .05 level). Although low, this measure indicates that there is indeed a positive low correlation between students' score in the previous module and their engagement in learning activities. An analysis of how tutors rate the students based on the observed learning activities will provide further clarity on the correlation between students' learning activities and previous module.

Overall, the findings from the three studies are that (1) the measures of the intra-class correlations are reasonably high, (2) the time-point of rating scale administration is a source of variance, (3) ratings on four the five problem characteristics are stable, and do not seem to be influenced by the students' competency in subject, achievements in previous test and learning activities. These findings generally support the reliability of the problem quality rating scale across different student groups. This is in line with a study by Schmdit, Marchais, Dolmans, and Gijselaers (1989) showing that students' ratings of PBL courses are generalizable across courses, between items, between different tutorial groups for several variables including perceived relevance of learning, quality of problems, group functioning, and tutor performance. However, their study was carried out at the general curricular level and not on individual problems as ours.

The implications of the findings are that the problem quality rating scale can be used to (1) assess students' experience of problem/s over time within the PBL cycle, and (2) assess the quality of problems. The results also indicate that it is preferable to assess the quality of problems after students have engaged in solving/tackling the problem. Variance in students' rating of problem interestingness indicates that the items encompassing this factor may need to be refined. It is possible that some of the items in the factor are not clear in terms of whether the reference is to the rater or the problem. For example, items like "I was not interested to read the problem" and "I was curious to find the answer" can be associated more with the self and personal interest than the problem. Hence these items could be refined as "The problem triggered my interest to read the problem", and "The problem triggered my curiosity".

There are limitations to the present study. One limitation is that only two problems are investigated. The results obtained may be different when tested with different problems. Future studies should test the rating scale with several problems. Furthermore, the use of the rating scale is retrospective and requires the students to have worked through the problems. The later limitation may be unavoidable as the items in the rating scale enquire about the students' learning activities as a result of working on the problem. Therefore, if the rater had not worked on the problem, then he/she would not be able to rate these items. Generlaizability of the findings to be further tested with students from different years and courses.

The findings of this study present itself for further exploration. Assuming that effective problems will lead to better learning, it can be postulated that a more effective problem will result in higher grades than a not-so-effective problem. To investigate which of the problem characteristics have contributed to the effectiveness, students can be asked to rate various problems thought to be differing in their effectiveness using the problem quality rating scale, and the ratings can be correlated with their academic achievements. This will provide further insights on the role of problems in PBL.

# CHAPTER 6

# DOES THE EXTENT OF PROBLEM FAMILIARITY INFLUENCE STUDENTS' LEARNING IN PROBLEM-BASED LEARNING?

## ABSTRACT

This study investigated the influence of problem familiarity on students' learning in problem-based education. To this end, two problems varying in the level of problem familiarity were presented to 172 students. Students' perceptions about their learning as a result of working on the problems, and tutor's assessment of the students' learning were collected at the end of learning activities. Results show that the students considered the familiar problem to be more interesting, and successful in guiding on the learning issues. However, they did not find the two problems to be significantly different in terms of the extent to which the problems stimulated critical reasoning and the extent to which the problems promoted collaborative learning. The tutors' assessment reflected the students' perceptions that the familiar problem had a more positive influence on their learning.

***Keywords:*** Problem-based learning, problem characteristics, problem familiarity, students' learning, students' perceptions, tutors' assessment

**INTRODUCTION**

Problem-based Learning (PBL) as an instructional approach is founded on the tenet that learning should be student-centered, self-directed, motivating, collaborative, and contextual (Hmelo-Silver, 2004). This draws a close parallel with cognitive psychological principles of learning (Norman & Schmidt, 1992). As in Vygotsky's learning theory (1978), the basic assumptions of PBL is that students bring with them a certain body of knowledge which serves as foundation to building more new knowledge through collaborative and constructive learning (Harland, 2003). In this approach, instructional material known as "problems" trigger the students' learning. When confronted with the problem, students in their small-groups work together to generate learning issues. The learning issues serve as guidelines for further exploration during their self-study. Findings made by the individual students during the self-study period are then shared with the group when they reconvene, and answers to the problem are co-constructed (Schmidt, 1983). Overall, the learning process in PBL is problem-driven and student-centered, resulting in a shift in the role of the tutor. Unlike teachers in traditional curriculum, tutors in PBL do not teach the students directly. Instead tutors assess the students' progress, and facilitate the students learning by stimulating discussions amongst team members, raising thought-provoking questions, encouraging collaborative work, and providing feedback at appropriate instances to the students (Das, Mpofu, Hasan, & Stewart, 2002; Maudsley, 1999).

To investigate the effectiveness of PBL, various studies have compared PBL curricula with traditional ones (Albanese & Mitchell, 1993; Colliver, 2000; Dochy, Segers, Van den Bossche, & Gijbels, 2003; Newman, 2003). Results from these studies are however contradictory. This has raised calls for more research, especially with a focus on understanding how and why PBL encourages students towards self-directed, active, collaborative and contextual learning (Dolmans, Gijselaers, Moust, De Grave, Wolfhagen, & Van der Vleuten, 2002; Norman & Schmidt, 2000). Taking the later approach, Schmidt and Gijselaers (1990) examined the interrelationship of the seven key elements of PBL; tutor's performance, problem quality, students' prior knowledge, time spent on self-study, group functioning, interest, and achievement using causal modelling. Their result showed that prior knowledge, quality of problems, and tutor performance influenced the amount of time spent on self-study activities and group functioning, which in turn influenced the students' interest and achievement (See also Gijselaers & Schmidt, 1990; Van Berkel & Schmidt, 2000). Extending on these studies, Van Berkel and Dolmans (2006) investigated the interaction between tutor performance, tutorial group productivity, and the effectiveness of a PBL course. They also found that the quality of problems influenced the students' achievement through group functioning. In addition, they found that the effectiveness of PBL problems is positively influenced by tutors' efforts to actively stimulate students' learning.

To gain more in-depth understanding of the PBL process, several studies have focused on the specific elements of PBL such as the attributes of effective tutoring (Schmidt, 1994; Schmidt & Moust, 1995), and collaborative learning (Dolmans & Schmidt, 2006; Slavin, 1996; Visschers-Pleijers, Dolmans, Wolfhagen, & Van der Vleuten, 2005). For instance, Schmidt and Moust (1995) found that effective tutoring is influenced by three distinct and interrelated qualities; tutors' content knowledge, tutors' willingness to become involved with the students in an authentic way and tutors' communication skills. Visschers-Pleijers, Dolmans, Wolfhagen, and Van der Vleuten, (2005) found that exploratory questions and cumulative reasoning factors explained 26% of the variance associated with group's productivity in PBL. Despite the wealth of knowledge we have gained about several aspects of PBL process, one critical aspect of the PBL process that has not received as much attention is the quality of problems (Hung, 2006).

In response to what determines the effectiveness of problems, it is essential to know the purpose of problems. Literature suggests that problems are purported to rekindle students' prior knowledge, spark discussions, encourage collaborative work, invoke problem-solving, promote self-directed learning, and lead to acquisition of relevant content knowledge (Hmelo-Silver, 2004). Existing theoretical guidelines and a limited number of empirical studies suggest what a good problem is (Des Marchais, 1999; Dolmans, Snellen-Balendong, Wolfhagen, & Van der Vleuten, 1997; Dolmans, Wolfhagen, & Scherpbier, 2003; Hung, 2006; Jacobs, Kim, Phillips, Pinsky, Brock, Phillips, & Keary, 2006; Shaw, 1976; Soppe, Schmidt, & Bruysten, 2005). Of these, Des Marchais' (1999) study is notable as his was the first study to provide an overview of nine attributes of problems using a systematic approach using expert's views.

To include the students' perceptions on effective problems, Sockalingam and Schmidt (2008) carried out focus group studies with eleven students and five tutors on characteristics of effective problems. They found eleven characteristics associated with effective problems. These characteristics are that problems should (1) be of suitable format (such as length of text and use of visuals), (2) be sufficiently clear, (3) lead to the intended learning issues, (4) be familiar to students, (5) be of appropriate difficulty level, (6) be applicable/relevant (for instance, to other modules/future work), (7) be interesting to the students, (8) promote self-directed learning, (9) stimulate critical reasoning, (10) encourage teamwork, and (11) trigger elaboration. No significant difference was found between the students' and tutors' views of effective problems. Overall, these characteristics were found to be in line with earlier mentioned studies. An interesting point is that three of these characteristics of problems are cited recurrently across the various literatures on characteristics of problems in PBL. These commonly cited characteristics are (1) the extent to which problem leads to intended learning issues, (2) the extent to which problem is familiar, and (3) the extent to which the problem is interesting. Even though the earlier mentioned literature suggest that the quality of problems in PBL is important for students' learning (Gijselaers & Schmidt, 1990; Schmidt, 1994; Van Berkel & Dolmans, 2006),and outline what makes a good problem, studies investigating how problems influence students' learning are not common.

One rare study is by Dolmans, Gijselaers, Schmidt, and Van der Meer (1993), who investigated the effectiveness of PBL problems. Their hypothesis was that an effective problem will lead students to the intended learning issues, in which case, there should be a match between the student-generated and faculty-intended learning issues. By comparing the learning issues generated by students with those intended by faculty for twelve problems, they found that students sometimes did not identify the faculty-intended learning issues. They attributed this to the complexity and unfamiliarity levels of the problems. Although this study was useful in identifying the effective problems, only one aspect of the problem, that is the extent to which the problem leads to the learning issues, was considered in detail.

Verkoeijen, Rikers, Te Winkel, and Van der Hurk, (2006) addressed this issue by taking an experimental approach to investigate the effect of goal clarity of problems on the quality and quantity of individual study. They presented one group of 60 students with a goal-free problem and another group of 60 students with a goal-specified problem. The two problems were similar in context and content except for the specification of a goal in the goal-specified problem. The authors supposed that the goal-free problem is likely to activate students' prior knowledge about more aspects of the problem, which will in turn encourage students to study more diverse information. They found their results to be supportive of their hypothesis as students in the goal-free condition read more articles, spent more time on self-study, and took more time to report the findings. An important facet of this study was that more information about the influence of problem on student's self study activities could be learnt. However, the conclusions were limited to student perceptions only.

Taking an experimental approach as well, Soppe et al. (2005) investigated the influence of problem familiarity on various aspects of students' learning. They, however, included external observations other than students' perceptions. To investigate the influence of problem familiarity, they presented 270 students from a psychology course with either a familiar or unfamiliar version of the same problem. Both the versions of this problem focused on the same subject/content matter (reasoning and decision-making); but they were considered to be differing in terms of familiarity level. The authors defined familiarity level as the extent to which the students could identify with the characters/actors in the story narrated in the problem. The familiar problem narrated events about psychology students. On the other hand, the unfamiliar version narrated events involving lawyers and other non-psychologists. Their hypothesis was that students working with the more familiar problem would activate more prior knowledge during the initial discussion. This activation of prior knowledge would stimulate more interest, which will in turn lead to students spending more time on self-study, resulting in acquisition of higher-quality subject matter which is reflected as higher scores on relevant knowledge tests.

To test their hypothesis, approximately half the 270 students received the familiar version of the problem while the other half received the unfamiliar version of the problem. Measurements of students' perceptions about problem familiarity, problem quality (interestingness, difficulty level and match with prior knowledge), time spent on self-study, and external measures such as tutor's assessment of the quality and quantity of learning issues, and students' achievement in knowledge tests were taken. Data analysis suggested that students found the familiar problem to be of higher quality and interesting than the unfamiliar problem. However, no corresponding difference was found in terms of the quality and quantity of learning issues generated by students, time taken for self-study, and students' achievement in knowledge tests as a result of the working on the two problems. The authors suggested that insignificant differences in some of the measures could be due to the subtleness of the experimental manipulation, and recommended increasing the difference between the familiar and unfamiliar version of the problem.

To address this, we relooked again at the definition of problem familiarity. Although we agree that problem familiarity includes the extent to which students can identify with the characters/actors in the story narrated in the problem, we felt that the notion of "familiarity" has a broader meaning. We define familiarity as the extent to which the problem matches the students' subject matter knowledge, experiential knowledge, and contextual knowledge, in accordance with Dochy and Alexandar's (1995) definition of prior knowledge. Like Soppe et al., (2005) we were motivated to investigate the influence of problem familiarity on students' learning, but this was based on the broader definition of problem familiarity. A caveat to note is that we did not attempt to differentiate between the various forms of prior knowledge associated with problem design. Resonating with Soppe et al.'s study (2005), our hypothesis was that the level of problem familiarity will influence students' learning, in particular their interest, critical reasoning, collaborative learning, and the extent to which students can identify the intended learning issues.

To test the hypothesis, we carried out an experimental study using two problems from the actual curriculum in a naturalistic educational setting. The two problems were identified *a priori* by the module coordinator to be differing in the extent of problem familiarity to students. The module coordinator selected the problems based on past experience in using the same problems. Of the two problems, one was on “Knowledge and morality” while the other was on “Realism and anti-realism”. Students were expected to be familiar with the problem on “Knowledge and morality” in terms of subject matter. In contrast, they were expected to be less familiar with the subject of “Realism and anti-realism”. The specific questions asked by the study are: (1) do students differentiate between familiar and unfamiliar problems?, (2) what are students’ perceptions of their own learning as a result of working on the familiar and unfamiliar problems, and (3) do the students’ perceptions of their own learning correspond with their tutors’ assessment?

To this end, we presented both the familiar and unfamiliar problem as part of the regular curriculum to 172 students on two occasions, one problem per occasion. A point to note is that the same group of students attempted both the problems on two occasions unlike Soppe et al.’s (2005) study in which each group of students attempted one of the two versions of the same problem. Students were administered the problem presupposed to be unfamiliar (Realism and anti-realism) first. After the students had worked on each of the two problems, students’ perceptions of problem familiarity, the extent to which the problem led to the intended learning issues, the extent to which problem triggered interest, the extent to which the problem stimulated critical reasoning, and the extent to which the problem promoted collaborative learning were collated using a validated rating scale (Sockalingam, Rotgans, & Schmidt, 2008). In addition, tutors' assessment of students' learning was used to verify the influence of the two problems.

## METHOD

### *Participants*

A total of 172 students in their first-year of studies in academic year 2007-2008 participated in the study. The students were in their second semester of PBL learning. The mean age of the entire sample was 18.64 years. Eighty-three of the participants were female (*M* = 18.11 years, *SD* = 1.00) and 89 of the participants were male (*M* = 18.51 years, *SD* = 1.78).

### *Educational context*

The research was conducted at Republic Polytechnic, Singapore. The institute 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). In this approach students are required 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 25, students are grouped in teams of five, and are guided by a 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 information collated, and exchange ideas with their teammates. Following this, the students and the tutor reconvene at a second meeting to discuss the overall progress. 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. Throughout the day, tutors observe and assess the students' learning.

***Materials***

*Problems*. Two problems from the "Cognitive Processes and Problem Solving" module were used in the study. The problems were identified by the module coordinator to be different in the extent of problem familiarity to students based on experience from the previous runs of the problems. The module coordinator identified problem "Realism and anti-realism" to be unfamiliar and problem "Knowledge and morality" to be familiar. The rationale given was the likelihood of the students being exposed to these subject matters in their pre-polytechnic years. A copy of both the problems is attached in the Appendix D for reference.

*Measures of students' perceptions of the effect of problems*. A 32-item rating scale developed and validated by Sockalingam, Rotgans, and Schmidt (2008) was used to measure the students' perceptions of problem familiarity and four aspects of their learning: (1) the extent to which the problem led to the formulation of intended learning issues, (2) the extent to which the problem triggered interest, (3) the extent to which the problem stimulated critical reasoning, and (4) the extent to which the problem promoted collaborative learning. In their validation study, Sockalingam, Rotgans, and Schmidt (2008) showed that the rating scale was reliable and valid.

A copy of the rating scale is provided in the appendix (See Appendix C). Students were required to respond on a 5-point Likert scale of 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree) for all items.

*Assessment of students' learning by tutor.* In the educational setting studied, tutors' formatively assess individual students' learning and award them an individual grade known as *daily grade*. Tutors observe the students throughout the learning process, taking several criteria into consideration. The criteria are: students' participation in discussion, teamwork, time and resource management, ability to collate relevant information, demonstration of reasoning skills, indications of self-monitoring, critical thinking, reflection, and evidence of understanding. They report daily grade on a 5-point performance scale; 0 (fail), 1 (conditional pass), 2 (acceptable), 3 (good), and 4 (excellent) for every student and for every problem. This measure is considered to be unobtrusive and natural as the formative assessment by tutors is part of the regular routine in the learning environment. It has been shown elsewhere that the daily-grade demonstrated high levels of reliability (Chai & 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 indicate high reliability and good predictive validity of this measure.

***Procedure***

The rating scale was administered electronically at the end of the day after students had worked on the problem. Students were informed to think about the problem that they had worked on for the day when responding to the rating scale. They were given fifteen minutes to complete the rating scale.

***Analysis***

Average of the students' response on the five problem characteristics was computed (Sockalingam, Rotgans, & Schmidt, 2008). Paired t-tests were used to analyze students' perceptions and tutors' assessment.

**RESULTS**

First, to answer the question on whether students differentiate between the familiar and unfamiliar problem, a paired t-test was carried out on the measure of problem familiarity. The mean values for problem "Realism and anti-realism" (*M* = 3.06, *SD* = .58) was significantly lesser than that for problem "Knowledge and morality" (*M* = 3.51, *SD* = .45), [*t*(171) = 10.32, *p* = .00]. This validates the difference in familiarity level of the two problems as identified by the module coordinator.

Second, to answer the question about the influence of problem familiarity on the extent to which the problem (1) led to the formulation of intended learning issues, (2) the extent to which problem triggered interest, (3) the extent to which the problem stimulated critical reasoning, and (4) the extent to which the problem promoted collaborative learning, paired t-tests were carried out. This assessment is based on students' perceptions. Significant differences were found between the two problems in the extent to which problem led to the identification of the intended learning issues [*t*(171) = 7.74, *p* = .00] and the extent to which the problem triggered interest [*t*(171) = 3.38, *p* = 00]. However, no significant difference was found between the two problems in the extent to which the problem stimulated critical reasoning [*t*(171) = -1.83, *p* = .07], and the extent to which the problem promoted collaborative learning [*t*(171) = .93, *p* = .36].

To evaluate the influence of problem familiarity on students' learning using an external measure, tutors' assessment of individual students' learning for the two problems were used. As a regular practice, tutors assess the students' learning based on observation of various criteria such as the students' participation in discussion, teamwork, time and resource management, ability to collate relevant information, demonstration of reasoning skills, indication of self-monitoring, critical thinking, reflection, and evidence of understanding and give students an individual grade for the respective problem. According to tutors' assessment, the two problems differed significantly in influencing students' learning [$t(171) = 2.04$, $p = .04$]. They assessed the students' learning to be better for the familiar problem ($M = 3.22$, $SD = .51$) than the unfamiliar problem ($M = 3.09$, $SD = .86$). The descriptive statistics, t-values and probability on the five aspects of problem and tutor's assessment are presented in Table 1.

**Table 1**

*Students' Perceptions about the Level of Problem Familiarity and the Effects of Problems, and Tutor's Assessment of Students' Learning*

| Measures | Unfamiliar Problem *M* *(SD)* | Familiar Problem *M* *(SD)* | *t* (*N* = 172) | *p* |
|---|---|---|---|---|
| The extent to which the problem was familiar to students | 3.06 (.58) | 3.51 (.45) | 10.32** | .00 |
| The extent to which problem led to the formulation of intended learning issues | 3.33 (.53) | 3.68 (.46) | 7.74** | .00 |
| The extent to which the problem triggered interest | 3.51 (.58) | 3.66 (.52) | 3.38** | .00 |
| The extent to which the problem stimulated critical reasoning | 3.88 (.45) | 3.82 (.42) | -1.83 | .07 |
| The extent to which the problem promoted collaborative learning | 3.86 (.51) | 3.90 (.45) | .93 | .36 |
| Tutors' assessment of student learning (Daily grade) | 3.09 (.86) | 3.22 (.51) | 2.04** | .04 |

** Significant at $p < .05$ level.

**DISCUSSION**

The present study aimed to investigate the influence of problem familiarity on students' learning in a natural educational setting. In particular, the questions asked are: (1) do students differentiate between familiar and unfamiliar problems?, (2) what are students' perceptions of their own learning as a result of working on the familiar and unfamiliar problems?, and (3) do the students' perceptions of their own learning correspond with their tutors' assessment? To this end, a familiar and an unfamiliar problem were presented to a total of 172 students. Measures of students' perceptions of problem familiarity and their learning were taken using a validated rating scale (Sockalingam et al., 2008). In addition, tutors' assessment of students' learning was used to counter check students' perceptions. Data analysis of the students' response revealed that they perceived the two problems to be significantly different with respect to problem familiarity. As presupposed, the problem on "Knowledge and morality" was found to be more familiar. In addition, the results suggest that the students perceived the familiar problem to be significantly more interesting, and more successful in guiding them on the learning issues. Interestingly, they did not find the two problems to be significantly different in stimulating critical reasoning, and in promoting collaborative learning. Tutors who observed and facilitated the participants in the problem solving process assessed that the students' individual learning to be significantly better for the familiar problem than for the unfamiliar problem. Overall, this study seemed to support the notion that problem familiarity positively influences the students' learning. The implications and limitations of the study are discussed further.

Students' perceptions of the two problems on the measure of problem familiarity show that they found the problem "Knowledge and morality" to be more familiar when compared with the problem on “Realism and anti-realism. This finding supports the presupposition about the extent of problem familiarity. Students’ response on the rating scale (Sockalingam et al., 2008) shows that they attributed this to a better fit of the problem with their prior knowledge associated with past-experience, subject knowledge, and general knowledge. As for the influence of problem familiarity on the measure of problem leading to the formulation of intended learning issues, students who worked on the familiar problem felt that it had clearer instruction, and had more key words or clues embedded to guide them successfully to the intended learning issues than the unfamiliar problem. This result reflects Dolmans et al.’s (1993) findings that students fail to identify learning issues intended by the tutor if their prior knowledge does not sufficiently match the problem. Students also felt that the familiar problem was significantly more appealing and engaging. This is in line with Soppe et al’s (2005) and Gijselaers and Schmidt’s (1990) findings that problems that familiar in context and problems that match students’ prior knowledge are found to be more interesting to students. The present study therefore adds further evidence that the extent of problem familiarity positively influences students’ interest and success in identifying relevant learning issues.

Surprisingly, at first glance, the results seem to suggest that there was no significant difference between the two problems in the extent to which critical reasoning was stimulated. Based on Hmelo-Silver’s (2004), description of the PBL process, it is expected that a more unfamiliar problem would trigger more questions and thinking. A closer examination of the subscales demystifies the anomalous findings. The rating scale used in the study (Sockalingam et al., 2008) defines critical reasoning as stimulation of questioning, thinking, and reasoning, and consideration of the problem from multiple perspectives. A closer examination of the students’ response shows that the unfamiliar problem had indeed resulted in significantly more questioning, and thinking and reasoning. However, no significant different emerged between the two problems in triggering consideration of multiple perspectives. This could be because both the problems contained goal specifications that required consideration of multiple perspectives (see Appendix D for the problems).

As for the influence of the problem familiarity on the collaborative learning, no significant difference was found between the two problems. Similar results have also been reported by Van Berkel and Schmidt (2000), and Gijselaers and Schmidt (1990) that prior knowledge does not necessarily influence collaborative learning. However, no explanation has been proposed for this. The result from this study suggests that unfamiliar problem stimulates significantly more questioning, thinking and reasoning than the familiar problem. Hence we think that exploring the type and pattern of verbal interaction may provide more insights than considering collaborative learning as a whole. This idea seems to be in line with, Visschers-Pleijers et al.'s (2005) study which showed that interaction in PBL can be classified into at least two types: exploratory questions and cumulative reasoning. However their study investigated interaction taking place as a result of students working on one problem only. Hence, a future study that investigates the influence of various types of problems on the pattern of group interaction may be useful.

Overall, the measures of students' perceptions suggest that the familiar problem has a more positive influence on their learning. This is also reflected in the tutors' assessment of the students' learning. In comparison with Soppe et al.'s (2005) study, the present study has not only considered more aspects of students' learning such as critical reasoning and collaborative learning, but it has also included the tutors' observation throughout the students' learning process which is expected to be more comprehensive than the measure of students' achievement through knowledge tests. In addition, this study has analyzed students' individual responses, rather than the aggregation of their responses at group level as in Soppe et al.'s (2005) study.

Nevertheless, the study has a number of limitations. The first limitation is the generalizability of the study. As only two problems were used in the study, the findings need to be tested with more problems, and other PBL contexts. Second, the study does not demonstrate causality between the various variables. A causal modeling approach may be more suitable for that purpose. The experimental/evaluative approach used in this study is instead more suitable to collate information about individual problems. This approach is likely to be useful as formative evaluation of the course material. Third, the study only focuses on the extent of problem familiarity. However, other confounding attributes of problems such as the level of problem clarity and problem complexity may be very closely associated with problem familiarity and this can influence the students' learning as well. Fourth, although this study seems to suggest that problem familiarity leads to better learning, one must be cautious in interpreting the result. It is likely that if a problem is too familiar, students may find it to be boring and not motivating enough, resulting in poor learning. Hence an extension of this study to investigate more problems will be useful.

Implications of the finding from this study for problem design are that students' prior knowledge must be carefully considered in designing effective problems. One strategy to designing familiar problem would be to use content and context that are familiar to students. However it may not be necessarily bad to have some elements of unfamiliarity incorporated in the problem as this is shown to stimulate thinking. A second strategy is to embed keywords and clues in the problem to help students in their problem analysis. In sum, various aspects of the problem such as content familiarity, context familiarity, embedding of keywords/clues in problem, and goal clarity may need to be considered in totality in designing effective problems. To extend this study’s findings, future research could include investigation into more problems that vary in degree of problem familiarity and also explore the type and pattern of verbal interactions taking place in groups.

# CHAPTER 7

## SUMMARY AND CONCLUSIONS

The research presented in this dissertation focused on the characteristics of problems in Problem-based Learning (PBL). Given the importance of the quality of problems, the challenges faced in designing problems, the lack of studies on characteristics of problems, and the shortcomings of the existing studies on characteristics of problem, I was motivated to find out more on the characteristics of problems. To this end, five studies were carried out using different methods in natural settings. Studies 1 and 2 were explorative in nature and investigated the students' and tutors' perceptions of characteristics associated with effective problems. Studies 3 and 4 developed a rating scale to measure the characteristics of problems, and tested the validity and reliability of the rating scale measures. Study 5 attempted to apply the rating scale in investigating the influence of problem familiarity on students' learning. Taken together, these five studies not only tried to shed light on the different characteristics associated with problem quality (Studies 1, 2, 3,& 4) but they also attempted to relate the characteristics to students' learning (Studies 2 & 5). These studies were carried out in two PBL settings.

Studies 1, 2, 3, and 5 were carried out at Republic Polytechnic in Singapore which uses PBL as the sole educational method while Study 4 was carried out at the Institute of Psychology, Erasmus University in The Netherlands, where PBL is the main educational method. The Polytechnic under study is the newest and fifth polytechnic in Singapore. This Polytechnic has been in operation since 2002 and has pioneered implementation of PBL as the sole educational method. The mission of polytechnics in Singapore is to equip students with knowledge and skills to prepare them for their future work. Typically, students join polytechnics post-secondary as their pre-university education.

The implementation of PBL in Republic Polytechnic in Singapore can be considered as unique in the sense that the PBL learning process takes place in one day; that is, students work on one problem each day (Alwis & O'Grady, 2002). In this one-day, one-problem implementation, students work in teams of five members, under the guidance of a tutor. Four to five teams make up a class. A typical day starts with the tutor introducing a problem to the students. An example of an explanation problem taken from the Cognitive Processes and Problem Solving Skills module is given below.

*Example of an Explanation Problem from a Cognitive Processes and Problem-solving Skills Module implemented in Republic Polytechnic*

**Education, what is it?**

Ivan Pavlov was a Russian biologist who received the Nobel Prize in 1904 for Medicine. He found out during a study that every time a bell is sounded when a dog is given food, the dog would salivate. Eventually, the dog would salivate even when just the bell rang without food.

Psychologists who had defined learning as what causes a “change in behaviour” concluded that the dog has learned something which it could not do before. This happening of “learning” in the dog has since become a famous example of “classical conditioning” in the so-called Learning Theory.

Sceptics criticize that if we link learning to change in behaviour, then if someone suffered a leg injury and started to limp, it would be acceptable to say that the injured person had learned to limp.

Quite clearly, there is so much confusion about learning. However,

the more important question to individuals, communities and tax-payers, is about education rather than learning. Some people believe that learning is the same as receiving an education, yet many would be unwilling to consider that the Pavlov's dog got educated to salivate, or someone got educated to limp following an injury.
What could be meant by the phrase "receiving an education"? What makes someone "educated"?

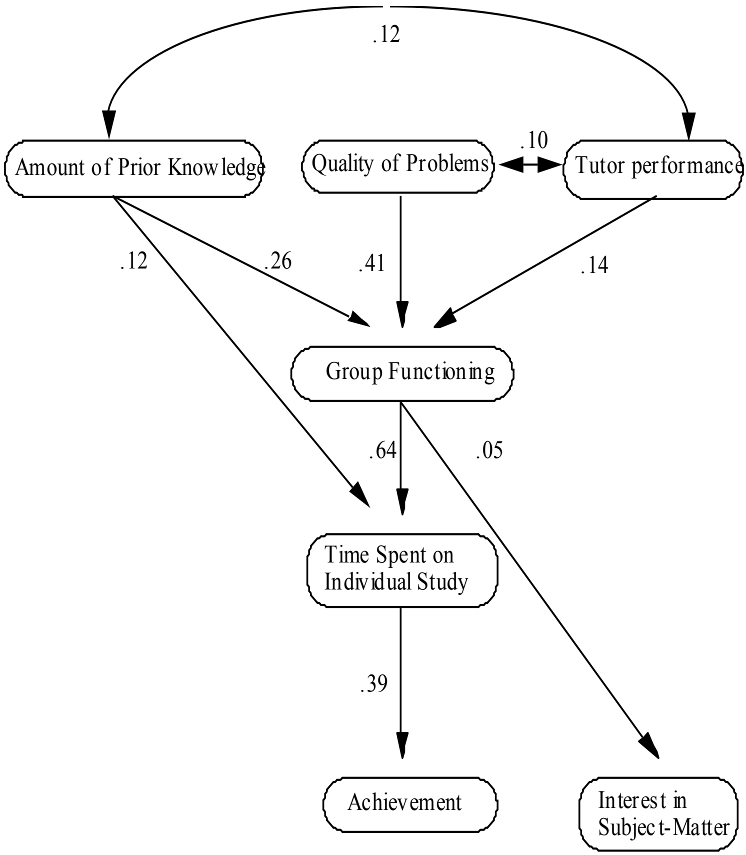
The problem serves as the starting point for students' learning process in PBL. Students begin working on the problem with no preliminary preparation but just their prior knowledge. They work in their collaborative teams to analyze the problem and try to explain the phenomenon described. In this process, they come up with a tentative hypothesis about the problem. For instance, using the earlier mentioned example, students may recognize from the problem that learning is considered to be more than responding to a stimulus. However, from their personal experience, they may associate learning with going to school to get an educational certificate. Or they may propose that the more information one knows, the more educated the person is. Yet another may counter this notion by hypothesizing that knowing more information may not mean that the person can apply this knowledge. Such discussion surface several questions which may lead to proposition of tentative conclusions. One example of a tentative conclusion could be that "learning may not necessarily mean knowing information".

These questions and propositions then serve as guidelines for the students' self-study. Not only that, the identification of gaps in their understanding engages them in their self-study. As a result, students refer to various resources such as internet, books, and news articles to find out more on these issues based on what they deem as relevant and important to respond to the problem. After the self-study period, the students reconvene with their team to share their findings, explain their views and synthesize a shared understanding about the issues presented in the problem. Thus, the problems serve to engage the students, 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 (Hmelo-Silver, 2004). Each day of problem-solving encompasses three meetings with two self-study period in between the meetings.

On the other hand, in the Institute of Psychology at Erasmus University, Rotterdam, the learning process takes place over a longer duration of one week and encompasses two meetings with a self-study period in between the meetings lasting two days. A second key difference is that the students work in larger groups of 8-10 members. Despite the differences in the implementation of PBL in the two institutions, the commonality is that both (1) use problems to initiate the learning process (with minimal lectures in the case of Erasmus University), (2) requiring students to work collaboratively, (3) as well as independently such that they carry out self-directed learning, (4) under the flexible guidance of a tutor, (5) ensuring that they have ample time for collaborative work and self-study. These characteristics are the hallmarks of PBL (Schmidt, Van der Molen, Te Winkel, & Wijnen, 2009).

Of these, the problems, students and tutors can be considered as the three “input” elements of PBL (Majoor, Schmidt, Snellen-Balendong, Moust, & Stalenhoef-Halling, 1990). In an investigation by Gijselaers and Schmidt (1990) on the interrelationship between the various elements of PBL, it was found that of these three “input” elements, the quality of problems had a more direct and stronger influence on the various “process” elements such as individual’s self-study time, group functioning and “outcome” elements such as achievement and interest than the other “two” input elements (see Figure 1). This study has been repeated and re confirmed by Van Berkel and Schmidt (2006) and Schmidt, Bruysten, and Soppe (2003). Figure 1 shows Schmidt el al.'s findings in a causal model.

**Figure 1**
*Causal Model of PBL*



What this means is that a good problem leads to improved learning. An implication of this result is that learning can be positively influenced by designing better problems. However, PBL practitioners face challenges in designing and evaluating problems (Angeli, 2002). While there are some existing studies that shed light on designing and evaluating problems (Des Marchais, 1999; Dolmans, Gijselaers, Schmidt, & Van der Meer, 1993; Jacobs, Dolmans, Wolfhagen, & Scherpbier, 2003; Marin-Campos, Mendoza-Morals, & Navarro-Hernandez, 2004; Mpofu, Das, Murdoch, & Lanphear, 1997; Soppe, Schmidt, & Bruysten, 2005; Verkoeijen, Rikers, Te Winkel, & Van der Hurk, 2006), there are six limitations to these studies.

First, not many studies use an empirical approach to define a broad range of problem characteristics. Second, studies that define problem characteristics include only the students' perceptions or tutor's perceptions but not both. Third, the relationship between problem characteristics and students' learning are not generally assessed by studies exploring a wider spectrum of problem characteristics. Fourth, some of these studies use resource and time-intensive methods, such as in the case of comparing the student-generated learning issues with the faculty-intended learning issues. Though such research intensive methods may provide detailed information, this may not be applicable when the quality of several problems need to be evaluated. Fifth, instruments used to measure the problems characteristics such as Jacobs et al.'s (2003) questionnaire and Marin-Campos et al.'s questionnaire (2005) focus on a selected few characteristics. In addition they are seldom validated (Soppe et al., 2003; Verkoeijen et al., 2006). Sixth, students' perceptions about the influence of problems on their learning are not corroborated with other measures.

These limitations raise several questions which can be summarized as follows:

- What are the students’ perspectives of effective problems?
- Do students consider several problem characteristics?
- Do students and tutors share a common understanding about these characteristics?
- Do students and tutors actually consider these characteristics when evaluating specific problems?
- Do students’ and tutors’ ratings of the problems correspond with the students’ grades?
- Is it possible to develop a rating scale to assess a more comprehensive list of problem characteristics than what is available at present?
- Is it possible to validate and test the reliability of such a rating scale?
- Can such a rating scale be used to assess the influence of problem characteristics on students’ learning?

These questions serve as the motivation behind the five studies carried out as part of this dissertation. The five studies could be found in Chapters 2 to 6 of this dissertation. The following sections presents the summary and conclusions drawn from the five studies in three broadly defined categories as (1) Characteristics of problems in students’ and tutors’ perceptions, (2) Development and testing of validity and reliability to measure the characteristics of problems, and (3) Application of the validated rating scale to examine the influence of problem familiarity on students’ learning.

## CHARACTERISTICS OF PROBLEMS IN STUDENTS’ AND TUTORS’ PERCEPTIONS

### STUDY 1

Study 1, reported in Chapter 2, aimed to capture panoramic ‘view’ of critical problem characteristics in students’ perceptions. Despite the existence of principles and guidelines which give list of problem characteristics, most of these tend to be theoretical and not validated. The few limited empirical studies tend to focus on specific characteristics. One rare study which has identified a comprehensive list of problem characteristics using Delphi technique is Des Marchais' study (1999). This study identified nine characteristics associated with effective problems. Though useful, this study had not included the students' perceptions. Students are the end-users of the problems and can be considered as novices in terms of content knowledge in comparison with the faculty. Cognitive psychology suggests that experts and novices process information differently (Van der Vleuten, 1996). Hence, it is possible that tutors' perceptions of the quality of problem are different from that of the students' perceptions. This motivated us to find out the students' perceptions on the characteristics associated with effective problems.

To this end, we asked 34 second year biomedical students from the polytechnic in Singapore to write a reflective essay on characteristics of good problems. The students who participated in this study were all in the second semester of the second year in the PBL curriculum. On average, each student had worked on over hundred problems. Students had each taken 4 modules per semester, consisting of 16 problems per module. Drawing on the participants' experience of solving this many PBL problems was considered to be useful in providing insight on the problems.

These essays were compiled and analyzed using TextSTAT text analysis software, obtained from the web link http://www.niederlandistik.fu-berlin.de/textstat/ (Huning, 2007). TextSTAT is a simple concordance program for the analysis of texts using data in the ASCII/ANSI/ HTML/ Microsoft Office format and is designed to count the word frequency in the input data. The program generated a frequency list of words in the students' responses. From the list generated, appropriate evaluative words associated with various qualitative aspects of problems were identified manually and categorized based on semantic similarity. The assumption was that the more often a particular characteristic was mentioned, the more important it was for students. This text analysis approach yielded eleven problem characteristics. Table 1 shows the ranking of the identified problem characteristics by importance. Of the eleven characteristics, the extent to which problem led to the intended learning issues was the most important characteristic to students, and the extent to which the problem stimulated discussion and teamwork were the least important characteristics to students.

The result from Study 1 shows that it is possible to capture a panoramic view of problem characteristics through students' perceptions. A comparison of the eleven characteristics from this study with the literature on problem quality (Des Marchais, 1999; Dolmans et al., 1997 Shaw, 1976;) shows that the students also referred to similar characteristics as those proposed in the literature. See Table 2 for the comparison of students' perspective from this study with other empirical studies (e.g., Des Marchais, 1999), and theoretical guidelines (e.g., Dolmans et al., 1997). This could be possibly because the students are constantly exposed to principles of constructivist learning as part of their PBL curricula. Hence, they may align their beliefs with the principles of constructivist learning that learning occurs as a result of engaging in self-directed learning as well as collaborative work to find solutions to authentic problems, which results in gain in their content knowledge, and interest (Savery & Duffy, 1995). Other studies have shown that students associate these principles in practice. For instance, Loyens, Rikers, and Schmidt, (2007) demonstrated that students recognize the distinctiveness of constructivist assumptions.

**Table 1**
*Key Characteristics of Problems and Ranking by Importance to Students*

| A problem should… | Words used by students | Frequency percentage of words used | Ranking of importance* |
|---|---|---|---|
| lead to learning issues | learn, issues, facts | 23..8 | 1 |
| trigger interest | interesting, like, capture | 11.5 | 2 |
| be of suitable format | phrase, picture, sentence | 10.9 | 3 |
| stimulate critical reasoning | thoughts, ideas, logic | 10.2 | 4 |
| promote self-directed learning | research, explore, tackle | 10.0 | 5 |
| be of suitable clarity | obvious, clear, understand | 7.3 | 6 |
| be of appropriate difficulty | easy, difficult, hard | 7.1 | 7 |
| enable application or use | apply, world, use | 7.0 | 8 |
| relate to prior knowledge | know, remember, background | 6.7 | 9 |
| stimulate elaboration | elaborate, brainstorm, discuss | 3.6 | 10 |
| promote teamwork | team, class, together | 1.9 | 11 |

*According to scale of importance from 1 to 11, 1 being the most important

Comparing the eleven characteristics from this study with Des Marchais' list of nine characteristics (1999), we can see that the students identified all of the nine problem characteristics cited by the experts. In addition, the students identified new problem characteristics such as problem format, problem difficulty, the extent to which the problem stimulates discussion, and promotes teamwork. More noticeably, the students differed from the experts in the ranking of the problem characteristics. For instance, the experts in Des Marchais' study (1999) identified the two most important criteria as (1) the extent to which the problem stimulates thinking/reasoning and (2) the extent to which the problem leads to self-directed learning in the students. However, the students in this study identified the extent to which the problem leads to intended learning issues as the most important characteristic. An explanation for the differences observed could be that the roles of the experts and students are different. Hence, their expectations of the quality of problems can be different. In line with this are studies which show discrepancies between the students' and tutors' perceptions of PBL. For instance, Gerzina, McLean, and Fairley (2005) showed that students and tutors differed significantly in their perceptions of the extent to which theoretical knowledge was applied in clinical settings; that is, more students than tutors perceived a link between the theory and application of it in the concerned dental clinical teaching program. In another study, Zanolli, Boshuizen, and De Grave (2002) showed that students and tutors differed in their ratings of several aspects of PBL. While the students ranked the tutors as the most important factor for their learning, the tutors ranked the students as the most important factor for the same. In addition, the students and tutors disagreed significantly on factors such as assessment and problem, with the students generally having a higher means than the tutors.

Another possibility could be that the objective and implementation of PBL curricula in Rouen University and the Polytechnic in this study are different (Des Marchais, 1999). On the one hand, Des Marchais' (1999) study was conducted in a medical university, while, the present study was conducted in a polytechnic which employs PBL curricula across all modules. Schmidt, et al., (2009) point out that the implementation of PBL can be varied depending on the objectives of PBL, and has proposed a categorization of the various versions of PBL into three types based on its objectives. Type I PBL focuses on information processing and is founded on the cognitive psychology principles of mental-model construction. Type II PBL is process-oriented, focusing on problem-solving skills such as clinical reasoning, and type III PBL focuses on learning skills which help students learn how to learn. As the PBL curriculum in Des Marchais' study (1999) is of medical context, it is possible that it focuses more on problem solving skills, and is of type II PBL. On the other hand, the institution involved in this study has adopted PBL across its curriculum and focuses more on knowledge construction as in type I PBL. Although reasonable, this postulation needs to be examined further. One way to overcome this difficulty in future studies will be to compare the perspectives of students and tutors from the same institution/the same type of PBL curricula. Understanding the difference in student-tutor perceptions will be important in interpreting program evaluation by the two groups.

Overall, this study attempts to present a wider spectrum of the characteristics of good problems in students' perspective. The students' responses pertaining to the various characteristics (See Appendix A) provides some insights into problem designing. For example, we can infer from the students' responses that the format of the problem can be modified to have an influence on students' interest in the problem. Context of the problem could be designed such that students are able to relate to some application or use in other modules or real-life context to have an impact on students' interest and learning. Appropriate key words and hints can be included in the problem to alter the clarity level of the problem so as to have an impact on the extent to which the problem leads to intended learning issues. The difficulty level of the problem could be adjusted to have an influence in the extent to which the problem is interesting, leading to learning issues, stimulates critical reasoning and promotes self-directed learning. Inferring from these responses, we propose that these eleven characteristics can be classified into two groups as either "features" or "functions", based on their roles. "Features" of the problems refer to characteristics that are design elements of the problems. Characteristics such as problem format, clarity, familiarity, difficulty and relevance (application and use) are such design elements of the problems.

On the other hand, "function" characteristics refer to the potential outcomes of engaging with or working on the problems. Of the eleven identified characteristics, the extent to which the problem stimulates critical reasoning, promotes self-directed learning, stimulates elaboration, promotes teamwork, stimulates interest, and leads to the intended learning issues are such function characteristics. In a way, these function characteristics are reflective of the five principles of constructivist learning and the objectives of PBL (Mayer, 1999; Savery & Duffy, 1995). The principles of constructivist learning are that learning occurs as a result of engaging in self-directed learning as well as collaborative work to find solutions to authentic problems, which results in gain in their content knowledge, and interest (Savery & Duffy, 1995). We propose that the "feature" characteristics of the problems could be modified to have an influence on the "function" characteristics. However, the results from this study also suggests that there exists a complex relationship between the feature and function characteristics. Hence, further research is needed to unravel this complex relationship to understand how to design effective problems.

As for problem evaluation, the results add further support to the existing understanding that the extent to which the problem leads to the intended learning issues is an important indicator of the problem effectiveness. In addition, the results add other characteristics that need to be considered in evaluating the effectiveness of problems. For instance, the function characteristics of the problems are likely to serve as appropriate indicators of problem effectiveness as these characteristics represent the objective of PBL. Therefore, measuring these characteristics could be used to indicate to what extent the problem plays a role in the effectiveness of PBL. To evaluate the effectiveness of problems, future studies could investigate how the feature characteristics influence the function characteristics of problems.

The limitations of this study are that first, it uses the students' essay responses to derive the characteristics of the problems for PBL. So there is a possibility that the students' vocabulary spectrum could limit the mention of words associated with certain problem characteristics resulting in neglect of these characteristics. Second, there is a possibility that some words can be categorized into one or more characteristics of problems. For instance, the word "long" could be associated with problem format as well as the extent to which the problem promoted self-directed learning as the word could be in reference to length of time. The implication of this is that the ranking of the importance of characteristics based on frequency count of the words may not be absolute. However, from the students' response, it is inferable that the students do consider the eleven characteristics. This also stresses the need for future studies to develop an instrument such as a rating scale to measure the characteristics of problems with higher reliability. Third, this study has only included the perspectives of a small group of students. Therefore, it will be better to explore the association of these problem characteristics with students' learning using a larger group of students and different approaches in the following studies. Fourth, this study did not include the tutor's perspective from the same institution or from an institution which adopts similar type of PBL (Schmidt et al., 2009). Fifth, the students were not given concrete sample problems to refer to. Hence the participants could have referred to different problems mentally.

**Table 2**

*Summary of the Characteristics of Problems for Problem-based Learning Stipulated by Various Studies*

| 11 criteria<br>This study<br><br>(Ranked by importance indicated by the prefix)<br><br>Problem should… | 8 criteria<br>Schmidt (1985)<br><br>(No ranking) | 9 criteria<br>(Des Marchais, 1999)<br><br>(Ranked by importance indicated by the prefix)<br><br>Problem should… | 7 principles<br>Dolmans et al. (1993)<br><br>(No ranking) | 5 features<br>Shaw (1976)<br><br>(No ranking) |
|---|---|---|---|---|
| 1. Lead to learning issues | Learning output<br><br>Openness | 5. Lead to discovery of learning objectives | Should match one or more of the faculty objectives | Solutions multiplicity |
| 2. Trigger interest | Intrinsic Interest | 6. Arouse curiosity and interest | Enhance students' interests | Intrinsic interest |
| 3. Be of suitable format | | | | |
| 4. Stimulate critical reasoning | | 1. Stimulate thinking, analysis and reasoning | | |
| 5. Promote self-directed learning | Time on task | 2. Initiate self-directed learning - | Stimulate self-directed learning | |
| 6. Be of suitable clarity | Concreteness | 9. Contain appropriate medical analytical | | |

| | | | | |
|---|---|---|---|---|
| | Goal clarity | vocabulary | | |
| 7. Be of appropriate difficulty | | | | Difficulty |
| 8. Enable application or use | | 7. Be on topics related to public health<br><br>8. Consider a global perspective | In context that is relevant to future work | |
| 9. Relate to prior knowledge | Familiarity / Prior knowledge | 3. Relate to previous basic knowledge<br><br>4. Be proposed in a realistic context / Relate to previous basic knowledge | Adapt well to students' prior knowledge<br><br>Relevant to basic concepts in the context of a clinical | Familiarity |
| 10. Stimulate elaboration | | | Contain several cues to stimulate students to elaborate | |
| 11. Promote teamwork | | | | Cooperation requirements |

**STUDY 2**

Study 2 aimed to address the limitations mentioned in Study 1. It further explored the problem characteristics. The key differences between these two studies are that in the latter, (1) a focus group study approach was used to identify characteristics associated with effective PBL problems, (2) both students' and tutors' perceptions were included, (3) students' and tutors' were presented with eight sample problems that they had worked on previously, and (4) students' and tutors' judgment about the effectiveness of the eight sample problems were correlated with students' achievement. In Study 2, participants were asked to describe about characteristics of "effective" problems instead of "good" problems as we felt that term "good" may be vague.

The focus group interviews were conducted with 11 students and 5 tutors. Each focus group consisted of two to three students or tutors, and the interviews were conducted in two phases. In the first phase of group-discussion, we sought the students' and tutors' generalized opinions about characteristics of effective PBL problems. In the subsequent phase, we gathered the students' and tutors' individual responses regarding the effectiveness of eight familiar sample problems from a Science module. Transcripts of the discussions from the first phase were text analyzed to identify the characteristics of effective problems in general. These characteristics were then used as criteria to frequency-score the students' and tutors' individual responses about the familiar problems. The resulting frequency scores were used to compare the students' and tutors' perceptions of problem characteristics associated with the effectiveness of the sample problems and to relate their perceptions with actual student grades for the subject matter covered by each problem.

The results from Study 2 suggest that both the students and tutors associated a total of eleven characteristics with the effectiveness of PBL problems in general. No new characteristics were found in this study when compared with Study 1. Both the groups identified that the most important characteristic of an effective problem is that it should lead to the intended learning issues. This characteristic is also considered by Dolmans et al. (1995) and Mpofu, Das, Murdoch, & Lanphear (1997) to be important in assessing the effectiveness of problems.

Findings from the second phase of the study on the eight sample problems showed that the eleven general characteristics of effective problems were considered by the students and tutors in judging specific problems as well. No new characteristics were generated by students and tutors when referring to specific problems. Implication of this result is that the eleven characteristics can be used to assess the effectiveness of specific problems as well as problems in general. Hence, we used the eleven characteristics as frequency-scoring criteria to frequency-score students' and tutors' responses with regard to the effectiveness of the sample problems.

To answer the 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. The 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, $\chi^2$ (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, $\chi^2$ (7, $N$ = 125) = 4.96, $p$ = .67. This indicated that the different roles played by the students and tutors in the students' learning process and the difference in their expertise did not have a significant influence on their judgments 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 learning progression in the students' learning process in PBL (Schmidt, 1983), the two groups could have noted similar elements influencing the students' learning. Kingsbury et al. (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. This result also suggests that it is more meaningful to compare the students and tutors working on the problem to provide feedback on the problems.

Finally, to answer the question of whether the eleven characteristics 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 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 students showed a high, significant and positive correlation, with an $r$ value of .75, $p <$

.05. This high and significant correlation between the tutor judgments and student grades suggest that the eleven characteristics 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 Processes 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 showed not only the repeatability of the study and confirmation of a high and significant correlation between the tutor judgment and student grades, but also the generalized use of the eleven characteristics 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 characteristics. Assessment of the effectiveness of sample problems using the eleven characteristics 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 characteristics are related to the effectiveness of the problems. Compared with other studies in the literature, this study seems to be the first to collate a list of characteristics 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 et al., 2008); but not both. This study also attempts to extend beyond identifying the characteristics by relating the eleven identified characteristics 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 characteristics of PBL problems do not relate the characteristics to the academic achievements.

Taken together, the results from Studies 1 and 2 suggest consistency in associating the eleven problem characteristics with effective problems. For instance, regardless of the different methods used in the two studies, different groups of student participants cited the same eleven characteristics. Results from Study 2 add to this that the tutors too identified the same eleven characteristics, both in general and when given specific problems. Comparison of the findings from Studies 1 and 2 with the Des Marchais study (1999), and theoretical guidelines/principles (Dolmans et al, 1997; Hung, 2003), reveal three common characteristics to all studies, and a new problem characteristic that has not been cited before. The three common characteristics are the extent to which the problem leads to intended learning issues, the extent to which the problem is interesting, and the extent to which the problem is familiar/related to students' prior knowledge. This commonality, regardless of the differences in the various studies could be taken to imply importance of these characteristics in problem design. The "new" problem characteristic identified is the format of the problem. Although literature on instructional design suggests that problem format needs to be considered in the design process and that problem format has an impact on students' learning (e.g., Hoffler & Leutner, 2007), this characteristic has not been explicitly included in the earlier mentioned studies. Hence, this dissertation brings up a relevant characteristic that may be overlooked by problem designers.

As far as we know, this study is also the first to evaluate the effectiveness of problems based on several characteristics and to have cross-validated the evaluation with students' achievement. The results suggest that the eleven characteristics seem to influence the students' achievement. However further research is needed to address how and why these characteristics influence students' achievement.

**MEASUREMENT OF PROBLEM QUALITY**
**STUDY 3**

Study 3 presented in Chapter 4 was a follow up of Studies 1 and 2, and aimed to develop a rating scale to measure the eleven problem characteristics. To this end, we first designed a 56-item rating scale to assess the eleven characteristics. The eleven characteristics are that problems should (1) be of suitable format (such as length of text and use of visuals), (2) be sufficiently clear, (3) lead to the intended learning issues, (4) be familiar to students, (5) be of appropriate difficulty level, (6) be applicable/relevant (for instance, to other modules/ future work), (7) be interesting to the students, (8) promote self-directed learning, (9) stimulate critical reasoning, (10) encourage teamwork, and (11) trigger elaboration. This rating scale was piloted with 185 first-year students from a polytechnic in Singapore. Confirmatory factor analysis showed the data did not adequately fit the hypothesized factor model. This is not uncommon in developing a new rating scale/questionnaire (Byrne, 2001).

We then analyzed the covariance matrix for items that did not contribute significantly to the underlying factors, or were highly correlated. Items that shared higher correlation with other factors, that is items which cross-loaded were combined to form a single factor, taking the conceptual validity into consideration. For instance, three of the characteristics; (1) suitable format of problem (such as length of text and use of visuals), (2) the extent to which the problem is clear, and (3) the extent to which the problem leads to formulation of intended learning issues were combined to form a single factor "the extent to which the problem leads to formulation of intended learning issues". Similarly, two other characteristics; (4) the extent to which problem promotes teamwork, and (5) the extent to which problem triggers elaboration were combined to form a single factor of "the extent to which the problem promotes collaborative learning". Next, items that did not contribute significantly to the underlying latent factor were dropped. This led to too few items for three of the characteristics. Given that initially these characteristics were only represented by four items, the three characteristics had to be excluded. The excluded characteristics were (6) the extent to which the problem promoted self-directed learning, (7) difficulty level of the problem, and (8) the extent to which the problem is applicable/useful. The remaining three characteristics of effective problems, (9) the extent to which the problem is familiar to students, (10) the extent to which the problem is interesting to students, and (11) the extent to which the problem stimulates critical reasoning, were considered to be unique and were used as individual factors in the rating scale.

This resulted in a 32-item rating scale, measuring five characteristics of the problems. The five factors of the rating scale are (1) the extent to which the problem leads to formulation of intended learning issues, (2) the extent to which the problem is familiar to students, (3) the extent to which the problem is interesting to students, (4) the extent to which the problem promotes collaborative learning, and (5) the extent to which the problem stimulates critical reasoning. For details of the items, see Appendix C. This rating scale was administered to another group of 517 first-year students. Subsequently, confirmatory factor analysis and reliability measures were carried out to examine the psychometric characteristics of the rating scale. For the analysis, the 32-items of the rating scale were parcelled, that is combined in groups of two or three based on semantic overlap (Bandalos & Finney, 2001; Little, Cunningham, Shahar, & Widaman, 2002). A total of 14 parcels were formed. Parcelling is a common measurement practice used in latent variable analysis. A parcel can be defined as the average of the two or three indicator items (Little et al., 2002). A detailed description of each of the 14 parcels, accompanied with the indicator items, is given in Appendix C.

The results showed a sufficient fit of the data with the hypothesized model. The Chi-square/*df* ratio for the main sample, ($N$ = 517), was 2.06, $p < .01$, RMSEA = .05 and CFI = .98. All factor loadings, ranging from .59 to .81, were statistically significant and thus contributed significantly to the respective latent variable. The coefficient $H$ values of the five factors were also found to be satisfactory; values ranged from .66 (Critical reasoning) to .78 (Collaborative learning), with an average of .75, indicating reasonable internal reliability of the five factors. Finally, use of split-half samples to test the model showed no significant difference in the factor loadings, signifying cross-validation of the hypothesised model. Collectively, these psychometric characteristics of the 32-item rating scale reveal that it can be used to adequately measure students' perceptions of these five principal characteristics of problems.

Comparison of the rating scale with the existing instruments (e.g., Jacobs et al., 2003; Marin-Campos et al., 2004) shows that the rating scale is more comprehensive in measuring four additional characteristics such as the extent to which the problem is related to students' prior knowledge, interest generated by the problem, collaborative learning promoted by the problem, and critical reasoning stimulated by the problem. An interesting point to take note is that four of the five characteristics measured by the rating scale reflect the principles of constructivist learning (Savery & Duffy, 1995). These four characteristics are; the extent to which the problem leads to learning issues, the extent to which the problem stimulates critical reasoning, the extent to which the problem promotes collaborative learning, and the extent to which the problem stimulates interest. According to the classification in Study 1, these are the function characteristics. The fifth characteristic, the extent to which the problem is related to students' prior knowledge, is a feature characteristic as classified in Study 1.

As such, one possibility of further research is the application of the rating scale in experimental studies in which feature characteristics such as problem familiarity or problem relevance are manipulated to study the effect on function characteristics. This approach is similar to Soppe et al.'s (2005) study on problem familiarity and Verkoeijen et al.'s (2006) study on problem goal clarity. Study 5 utilizes the rating scale in the suggested manner. An advantage of this rating scale is that it can be used to investigate the impact of the problem on the various measures of students' learning as represented by the function characteristics. Given that the objective of PBL is broader than the traditional curriculum (Barrows & Tamblyn, 1980; Hmelo-Silver, 2004), use of conventional knowledge tests is not likely to represent the effectiveness of PBL (Hmelo-Silver, 2004). Hence, utility of the rating scale to study the various measures of the learning outcomes can be considered beneficial.

**STUDY 4**

Although Study 3 had shown that the measures of the five problem characteristics of problems are reliable in terms of internal consistency, it does not indicate whether these measures were stable over time and across different student groups. According to Cohen, Manion, and Morrison (2000) the three principles of reliability are internal consistency, stability, and equivalence. Internal consistency measure of the rating scale indicates the extent of congruency and consistency between the different items of a single scale. On the other hand, stability refers to the measure of consistency over time and across different raters. It is possible that the rating scale is reliable in terms of internal consistency, but only when measured at a certain time-point or when administered to a specific group of students. For instance, high achievers may consistently evaluate the problems differently from the low achievers.

Hence, the objective of this study was to assess the inter-rater reliability and the stability of the ratings over time, and across different groups of students in using the problem quality rating scale. More specifically, this research set out to investigate (1) the intra-class correlation of the students' ratings of two problems, (2) the temporal stability of the ratings across various time points during the learning cycle, and, (3) the stability of ratings across different ability groups of students. The intra-class correlation measures the agreement between the raters, in our case, on the characteristics of the problem. A value of "0" is considered as no agreement between the raters while the value of "1" is indicative of absolute agreement between the raters.

In testing the temporal stability in students' ratings over the period of the PBL learning cycle, one hypothesis is that the ratings are not influenced by the time-point of rating scale administration as students are responding on the same problem. If this is true, it means that students do not let their judgment of the quality of problem be influenced by how much experience they have with the problem. However, it is possible that students' experience with the problem has an impact on their ratings of the quality of the problem. For instance, students who have just read the problem may not find the problem to be as interesting as after they have discussed with their teammates. Several studies show that group functioning has a positive and significant influence on level of interest (Dolmans & Schmidt, 2006; Gijselaers & Schmidt, 1990; Schmidt & Gijselaers, 1990; Schmidt & Moust, 2000). Hence, students' perception of the extent to which the problem is interesting may be higher after the initial discussion than after reading the problem. It is also possible that students may report lower interest after the final discussion if their epistemic curiosity is satisfied (Schmidt, 1993). Likewise variations may also be observed in students' ratings of the other problem characteristics.

In testing the inter-rater reliability, we wanted to know if students' achievements in knowledge tests and their learning activities while working on the problem influenced how they rated the problem. It is possible that students who scored high in a previous module feel more positive/confident, and rate the subsequent problem to be good. It could also be that students who had not received good scores rate the subsequent problem to be bad. Studies have shown that students' grades influence their evaluation of course work and teaching (Brown, 1976; El Ansari & Oskrochi, 2004; Engdahl, Keating, & Perrachione 1993; Feldman, 1976). On the other hand, even if students do not know their grades, it is possible that high achievers and low achievers rate the problems differently. Boud and Falchikov (1989) reported that high achievers were more competent than low achievers in self-assessment. This motivated us to find out if students' achievement in knowledge tests reflected how they rated a given problem.

As the knowledge tests may not be absolutely representative of what happens during the learning process, we also wanted to investigate whether students exhibiting different level of learning activities as observed by the tutors perceived the problem quality differently. For instance, it may be possible that students who demonstrated high level of collaborative work rate the problem to be higher on the problem characteristics of "the extent to which the problem promoted collaborative learning". Dolmans and Wolfhagen (2005) reported that students' perceptions of their group productivity are highly correlated with their perceptions of the effectiveness of their learning unit. However, in their study, effectiveness of learning unit was considered at the general level and not on specific characteristics of the problem, and the measurement of students' learning activities were students' perceptions rather than tutor's observations. Hence we wanted to know if students observed to have demonstrated different levels of learning activities varied in their ratings of specific problem characteristics.

To this end, students were asked to rate two problems from the Personality Psychology module in Erasmus University using the problem quality rating scale. The two problems were (1) Genetic and biochemical roots of personality and (2) (In) stability of human behaviour. For the first problem, the rating scale was administered at one of the three time-points; after reading the problem, after the initial discussion, and after the final discussion. Students who were asked to respond after reading the problem were presented with a shorter version of the rating scale. As we felt that certain items in the rating scale such as "The problem captivated my attention throughout the day", "Our team worked efficiently" can only be answered after working on the problem, we excluded these items from the rating scale, resulting in a shorter version of 16-items. The shorter rating scale measured three of the five problem characteristics; the extent to which the problem leads to the learning issues, the extent to which the problem is familiar, and the extent to which the problem is interesting. The items used in the shorter version are indicated in Appendix C. As for the second problem, the rating scale was administered at one of the two time-points; after reading the problem, and after the initial discussion. As before, students responding after reading the problem used the shorter version and students responding after the initial discussion used the longer version of the rating scale. Students took 10-15 minutes to respond on the rating scale. They were not informed how the data was going to be analyzed. Average measures of the problem characteristics were computed at the individual student level.

First, intra-class correlation measure was computed across the different raters for each of the five problem characteristics for the two problems. Average of the intra class correlation for the five characteristics on the first problem (Genetic and biochemical roots of personality) was found to be .87. For the second problem ((In) stability of human behaviour), it was found to be .92, indicating the students had agreed on the characteristics important to problem quality. Following that, a series of multivariate analysis of variance was carried out to determine the temporal stability of the ratings across various time points during the learning cycle, and the stability of ratings across different ability groups of students.

The results for the temporal stability of the students' ratings studied at various points in time showed that the students rated the first problem to have led to more learning issues after the final discussion ($M$ = 3.61, $SD$ = .57) than after the initial discussion ($M$ = 3.12, $SD$ = .79). On the other hand, they felt that the problem had resulted in lower level of critical reasoning after the final discussion ($M$ = 3.01, $SD$ = .35) than after the initial discussion ($M$ = 3.17, $SD$ = .32).

This could be possibly because students' perceptions about their learning changes with time. For instance, students may brainstorm and come up with several unanswered queries during the initial discussion. Typically, these unresolved queries help to direct the students in their self-directed learning activities (Hmelo-Silver, 2004). Hence, after the initial discussion they may feel that the problem has not yet led to sufficient learning issues. However, by the end of the final discussion, students may feel that the problem has led to satisfactory level of learning issues as most of their queries are answered through self-directed and collaborative learning (Hmelo-Silver, 2004). This may also explain why students perceived the problem to have stimulated more critical reasoning after the initial discussion than after the final discussion. Students' queries may have been answered resulting in satisfaction of their epistemic curiosity by the end of the final discussion (Schmidt, 1993).

This idea is supported by other studies in literature. For instance, a study by Dolmans, Schmidt, and Gijselaers (1995) shows that students identified 64% of the faculty-intended learning issues by the end of the learning session. Another finding was that the learning issues generated during group discussion were not the only guiding force directing the students in their choice of what to study. They found that the students may decide to study on topics not directed by the learning issues. This means that what students had learnt right after reading the problem or initial discussion is likely to be different from what they learn by the end of the final discussion. Nonetheless, students would have learnt more by the final discussion.

Similar results were also found on the shorter rating scale. Interestingly, students' ratings on the extent to which the problem led to the learning issues was found to be higher after reading the problem ($M$ = 3.31, $SD$ = .59) than after the initial discussion ($M$ =3.11, $SD$=.32). This could be attributed to students' (mis) conceptions of what they thought the problem was about right after reading the problem. However, the initial discussion could have revealed several areas of knowledge gaps, resulting in the lowering of their ratings on the extent to which the problem leads to the learning issues after the initial discussion.

Although one would expect that more unresolved queries and perceived high level of critical reasoning would have corresponded with an increase in the interest level, the interest level was found to be unaffected over time for this problem. Students reported similar level of interest, after the initial discussion (*M* = 3.31, *SD* = .82), and after the final discussion (*M* = 3.39, *SD* = .59). As for students' ratings on the extent to which the problem was found to be familiar (after the initial discussion; *M* = 2.77, *SD* = .70, and after the final discussion; *M* = 2.91, *SD* = .60), and to have promoted collaborative learning (after the initial discussion; *M* = 3.19, *SD* = .84, and after the final discussion; *M* = 3.20, *SD* = .59), there was no significant difference between the two time points.

In contrast, ratings on the second problem showed (1) no significant difference in the ratings on the extent to which the problem led to the learning issues (after reading the problem; *M* = 3.11, *SD* = .65, and after the initial discussion; *M* = 3.04, *SD* = .68 ), and (2) significant differences in the extent to which the problem was found to be interesting (after reading the problem; *M* = 3.67, *SD* = .75, and after the initial discussion; *M* = 3.44, *SD* = .72 ), and familiar (after reading the problem; *M* = 2.90, *SD* = .53, and after the initial discussion; *M* = 2.76, *SD* = .47 ).These differences in the ratings of the two problems over time lead to three inferences. First, the results reveal that time-point of rating scale administration is a potential source of variance in students' ratings of the problem quality. The ratings seem to be the lowest after the initial discussion than after reading of the problem and after the final discussion for both the problems. Nevertheless, generalizability of this observation needs to be tested further. The results could also mean that students' experience with the problem is critical to their assessment of the problem quality. Most studies assessing the quality of problems do so after students have worked on the problem (Jacobs et al., 2003; Marin-Campos et al., 2004; Soppe et al., 2005; Verkoeijen et al., 2006). As far as we know, the present study may be the first exploration of whether the time point of problem quality assessment affects the students' ratings. The finding that students' ratings vary over time during the PBL cycle provides new insights. The results support the notion that engagement in problem is needed for assessment of the problem quality.

Second, these results also support the validity of the rating scale. Studies testing for reliability of ratings aim to show high reliability values. Conversely, consistent high values regardless of context may mean that the measure is invalid. For instance, students are generally expected to have learnt more at the end of the final discussion compared to the initial discussion. If the rating scale to assess problem quality is not able to discern this difference, then there is a possibility that the rating scale is not valid. The results from the first two studies indicate that the rating scale is able to detect the variations, therefore adding support to its validity.

Third, variation in problems seems to influence students' ratings. This is in line with studies by Marin-Campos et al., 2004; Soppe et al., 2005; Verkoeijen et al., 2006. However, it is not the objective of this research to decipher how exactly the quality of problem influences the students' ratings from this study. This presents itself for further research.

As for the stability of ratings across different ability groups of students, regardless of the scores obtained in module under study, students rated the problem similarly. A possible reason for the invariance of ratings could be that the students in our study are able to assess the problem quality equally well. The results also show that students who scored higher in the previous module rated the problem to be more interesting. This could be because these students were aware of their grades. Literature suggests that the students' rating of curriculum could be influenced by their grades (Brown, 1976; El Ansari & Oskrochi, 2004; Engdahl, Keating, & Perrachione 1993; Feldman, 1976).

Students who were observed to have demonstrated higher learning activities also rated the problem to be more interesting.

From literature, we know that task engagement (that is demonstration of learning activities) is known to be correlated with interest in learning activities (Krapp, Hidi, & Renninger, 1992). Interest in learning can be distinguished as due to personal interest and situational interest. Personal interests are specific to individuals and tend to endure and evolve over time and across different situations (Renninger, 2000). Situational interests refer to interests that are evoked by the interestingness of a situation or context (Krapp et al., 1992). In our case, we can define the interest triggered by the problem to be situational interest. As the three groups of students who were identified to be demonstrating different levels of learning activities were presented with the same problem, we can assume that the situational interest is not varied. Hence it is logical to infer that the students' personal interest in the problem could have led them to varying extent of learning activities. In that case, it seems reasonable that students observed to be highly engaged rated the problem as highly interesting (Krapp et al., 1992). What strikes is that students rated to be "low" in learning activities also rated the problem to be highly interesting. One possibility is that tutors could have rated the students' learning activities with reference to the faculty-intended learning issues. Studies show that often students explore topics that may be related but not necessarily directed at answering the problem (Dolmans et al., 1993; Dolmans et al., 1995; Mpofu et al., 1997). Hence such students may have been rated "low" by tutors.

In explaining the difference in students' ratings of problem interestingness, it is also possible that students' grades in the previous module boosts their confidence, resulting in a higher perceived ability (Parsons, Croft, & Harrison, 2009). Greene and Miller (1996) show that perceived ability is positively correlated with meaningful cognitive engagement. Cognitive engagement on a given problem in turn could reflect the level of interestingness that the students associate with the problem (Krapp et al., 1992). A correlation of the tutor's score of students' learning activities and their score in the previous module was found to be .14 (correlation is significant at .05 level). Although low, this measure indicates that there is indeed a positive low correlation between students' score in the previous module and their engagement in learning activities. An analysis of how tutors rate the students based on the observed learning activities will provide further clarity on the correlation between students' learning activities and previous module.

Overall, the findings from the three parts of Study 4 are that (1) the measures of the intra-class correlations are reasonably high, (2) the time-point of rating scale administration is a source of invariance, (3) ratings on four of the five problem characteristics are stable, and do not seem to be influenced by the students' competency in subject, achievements in previous test and learning activities. The results indicate that students' ratings of problem interestingness may need to be investigated further. The findings generally support the reliability of the problem quality rating scale across different student groups. This is in line with a study by Schmidt, Marchais, Dolmans, and Gijselaers (1989) showing that students' ratings of PBL courses are generalizable across courses, between items, between different tutorial groups for several variables including perceived relevance of learning, quality of problems, group functioning, and tutor performance. However, their study was carried out at the general curricular level and not on individual problems as ours.

The implications of the findings are that the problem quality rating scale can be used to (1) assess students' experience of problem/s over time within the PBL cycle, and (2) assess the quality of problems. The results also indicate that it is preferable to assess the quality of problems after students have engaged in solving/tackling the problem. Variances in students' rating of problem interestingness suggest that the items encompassing this factor may need to be refined. It is possible that some of the items in the factor are not clear in terms of whether the reference is to the rater or the problem. For example, items like “I was not interested to read the problem” and “I was curious to find the answer” can be associated more with the self and personal interest than the problem. Hence these items could be refined as “The problem triggered my interest to read the problem”, and “The problem triggered my curiosity”.

The findings of this study present itself for further exploration. Assuming that effective problems will lead to better learning, it can be postulated that a more effective problem will result in higher grades than a not-so-effective problem. To investigate which of the problem characteristics have contributed to the effectiveness, students can be asked to rate various problems thought to be differing in their effectiveness using the problem quality rating scale, and the ratings can be correlated with their learning. This approach is taken in Study 5.

## APPLICATION OF PROBLEM QUALITY RATING SCALE STUDY 5

Study 5 utilized the rating scale from Study 3 to investigate the influence of problem familiarity on students’ learning. This study can be considered as an extension of Soppe et al.’s (2005) study which investigated the influence of problem familiarity on various aspects of students’ learning. To do so, they presented 270 students from a psychology course with either a familiar or unfamiliar version of the same problem. The authors defined familiarity level as the extent to which the students could identify with the characters/actors in the story narrated in the problem. Both the versions of this problem focused on the same subject/content matter (reasoning and decision-making). The familiar problem narrated events about psychology students. On the other hand, the unfamiliar version narrated events involving lawyers and other non-psychologists. Their hypothesis was that students working with the more familiar problem would activate more prior knowledge during the initial discussion. This activation of prior knowledge would stimulate more interest, which will in turn lead to students spending more time on self-study, resulting in acquisition of higher-quality subject matter which is reflected as higher scores on relevant knowledge tests.

Approximately half the 270 students received the familiar version while the other half received the unfamiliar version of problem. Measurements of students' perceptions about problem familiarity, problem quality (interestingness, difficulty level and match with prior knowledge), and time spent on self-study were measured. In addition, external measures such as tutor's assessment of the quality and quantity of learning issues, as well as students' achievement in knowledge tests were taken. Data analysis suggested that students found the familiar problem to be of higher quality and to be more interesting than the unfamiliar problem. However, no corresponding difference was found in terms of the quality and quantity of learning issues generated by students, time taken for self-study, and students' achievement in knowledge tests as a result of the working on the two problems. The authors suggested that insignificant differences in some of the measures could be due to the subtleness of the experimental manipulation, and recommended increasing the difference between the familiar and unfamiliar version of the problem.

To address this, we relooked again at the definition of problem familiarity. Although we agree that problem familiarity includes the extent to which students can identify with the characters/actors in the story narrated in the problem, we felt that the notion of "familiarity" has a broader meaning. We define familiarity as the extent to which the problem matches the students' subject matter knowledge, experiential knowledge, and contextual knowledge, in accordance with Dochy and Alexandar's (1995) definition of prior knowledge. Like Soppe et al., (2005) we aimed to investigate the influence of problem familiarity on students' learning, but this was based on the broader definition of problem familiarity. A caveat to note is that we did not attempt to differentiate between the various forms of prior knowledge associated with problem design. Resonating with Soppe et al.'s study (2005), our hypothesis was that the level of problem familiarity will influence students' learning, in particular their interest, critical reasoning, collaborative learning, and the extent to which students can identify the intended learning issues.

To test the hypothesis, we carried out an experimental study using two problems from the actual curriculum in a naturalistic educational setting. The two problems were identified *a priori* by the module coordinator to be differing in the extent of problem familiarity to students. The module coordinator selected the problems based on past experience in using the same problems. Of the two problems, one was on "Knowledge and morality" while the other was on "Realism and anti-realism". Students were expected to be familiar with the problem on "Knowledge and morality" in terms of subject matter. In contrast, they were expected to be less familiar with the subject of "Realism and anti-realism". The specific questions asked by the study are (1) do students differentiate between familiar and unfamiliar problems?, (2) what are students' perceptions of their own learning as a result of working on the familiar and unfamiliar problems, and (3) does the students' perceptions of their own learning correspond with their tutors' assessment?

To this end, we presented both the familiar and unfamiliar problem as part of the regular curriculum to 172 students on two occasions, one problem per occasion. A point to note is that the same group of students attempted both the problems on two occasions unlike Soppe et al.'s (2005) study in which each group of students attempted one of the two versions of the same problem. Students were administered the problem presupposed to be unfamiliar (Realism and anti-realism) first. After the students had worked on each of the two problems, students' perceptions of problem familiarity, the extent to which the problem led to the intended learning issues, the extent to which problem triggered interest, the extent to which the problem stimulated critical reasoning, and the extent to which the problem promoted collaborative learning were collated using the rating scale from Study 3. In addition, tutors' assessment of students' learning was used to verify the influence of the two problems.

Data analysis of the students' response revealed that they perceived the two problems to be significantly different with respect to problem familiarity. As presupposed, the problem on "Knowledge and morality" was found to be more familiar. In addition, the results suggest that the students perceived the familiar problem to be significantly more interesting, and more successful in guiding them on the learning issues. Interestingly, they did not find the two problems to be significantly different in stimulating critical reasoning, and in promoting collaborative learning. Tutors who observed and facilitated the participants in the problem solving process assessed that the students' individual learning to be significantly better for the familiar problem than for the unfamiliar problem.

Students' perceptions of the two problems on the measure of problem familiarity show that they found the problem "Knowledge and morality" to be more familiar ($M$ = 3.51, $SD$ = *.45)* when compared with the problem on "Realism and anti-realism ($M$ = 3.06, $SD$ = *.58)*. This finding supports the presupposition about the extent of problem familiarity. Students' response on the rating scale (Sockalingam et al., 2008) shows that they attributed this to a better fit of the problem with their prior knowledge associated with past-experience, subject knowledge, and general knowledge. As for the influence of problem familiarity on the measure of problem leading to the formulation of intended learning issues, students who worked on the familiar problem felt that it had clearer instruction, and had more key words or clues embedded to guide them successfully to the intended learning issues (*M* = 3.68, *SD* = *.53)* than the unfamiliar problem (*M* = 3.33, *SD* = *.46)*. This result reflects Dolmans et al.'s (1993) findings that students fail to identify learning issues intended by the tutor if their prior knowledge does not sufficiently match the problem.

Students also felt that the familiar problem (*M* = 3.66, *SD* = *.52)* was significantly more appealing and engaging than the unfamiliar problem (*M* = 3.51, *SD* = *.58)*. This is in line with Soppe et al'.s (2005) and Gijselaers & Schmidt's (1990) findings that problems that are familiar in context and problems that match students' prior knowledge are found to be more interesting to students. The present study therefore adds further evidence that the extent of problem familiarity positively influences students' interest and success in identifying relevant learning issues.

Surprisingly, at first glance, the results seem to suggest that there was no significant difference between the two problems in the extent to which critical reasoning was stimulated (Familiar problem; *M* = 3.82, *SD* = *.42,* Unfamiliar problem; *M* = 3.88, *SD* = *.45).* Based on Hmelo-Silver's (2004), description of the PBL process, it is expected that a more unfamiliar problem would trigger more questions and thinking. A closer examination of the subscales demystifies the anomalous findings. The rating scale used in the study (Sockalingam et al., 2008) defines critical reasoning as stimulation of questioning, thinking, and reasoning, and consideration of the problem from multiple perspectives. A closer examination of the students' response shows that the unfamiliar problem had indeed resulted in significantly more questioning, and thinking and reasoning. However, no significant different emerged between the two problems in triggering consideration of multiple perspectives. This could be because both the problems contained goal specifications that required consideration of multiple perspectives.

As for the influence of the problem familiarity on the collaborative learning, no significant difference was found between the two problems (Familiar problem; $M$ = 3.90, $SD$ = *.45,* Unfamiliar problem; $M$ = 3.86, $SD$ = *.51).* Similar results have also been reported by Van Berkel and Schmidt (2000), and Gijselaers and Schmidt (1990) that prior knowledge does not necessarily influence collaborative learning. However, no explanation has been proposed for this. The result from this study suggests that unfamiliar problem stimulates significantly more questioning, thinking and reasoning than the familiar problem. Hence, we think that exploring the type and pattern of verbal interaction may provide more insights than considering collaborative learning as a whole. This idea seems to be in line with, Visschers-Pleijers et al.'s (2005) study which showed that interaction in PBL can be classified into at least two types; exploratory questions and cumulative reasoning. However, their study investigated interaction taking place as a result of students working on one problem only. Hence, a future study that investigates the influence of various types of problems on the pattern of group interaction may be useful.

Overall, the measures of students' perceptions suggest that the familiar problem has a more positive influence on their learning. This is also reflected in the tutors' assessment of the students' learning. According to tutors' assessment, the two problems differed significantly in influencing students' learning [$t(171) = 2.04$, $p$ =

.04]. They assessed the students' learning to be better for the familiar problem (*M* = 3.22, *SD* = .51) than the unfamiliar problem (*M* = 3.09, *SD* = .86). In comparison with Soppe et al.'s (2005) study, the present study has not only considered more aspects of students' learning such as critical reasoning and collaborative learning, but it has also included the tutors' observation throughout the students' learning process which is expected to be more comprehensive than the measure of students' achievement through knowledge tests. In addition, this study has analyzed students' individual responses, rather than the aggregation of their responses at group level as in Soppe et al.'s (2005) study.

Nevertheless, the study has a number of limitations. The first limitation is the generalizability of the study. As only two problems were used in the study, the findings need to be tested with more problems, and other PBL contexts. Second, the study does not demonstrate causality between the various variables. A causal modeling approach may be more suitable for that purpose. The experimental/evaluative approach used in this study is instead more suitable to collate information about individual problems. This approach is likely to be useful as formative evaluation of the course material. Third, the study only focuses on the extent of problem familiarity. However, other confounding characteristics of problems such as the level of problem clarity and problem complexity may be very closely associated with problem familiarity and this can influence the students' learning as well. Fourth, although this study seems to suggest that problem familiarity leads to better learning, one must be cautious in interpreting the result. It is likely that if a problem is too familiar, students may find it to be boring and not motivating enough, resulting in poor learning. Hence an extension of this study to investigate more problems will be useful.

**CONCLUSION**

In conclusion, the five studies in the dissertation build on what is found in the literature, and one another to synthesize new knowledge about problems used in Problem-based Learning. The first two studies in the dissertation capture a panoramic view of students' perceptions, and provide a comparison between both students' and tutors' perceptions for the first time. Results from these studies also identify three common characteristics cited across various PBL contexts and a new characteristic which has been neglected in the existing guidelines/principles of problem design. In addition to generating a comprehensive list of problem characteristics, Study 1 attempts to provide a meaningful classification of the characteristics as feature and function characteristics. Study 2 shows that there is no significant differences between the students' and tutor's perceptions of effective problems, and relates the effectiveness of problems in terms of the eleven identified characteristics to students' learning. Overall, Studies 1 and 2 form the foundations of the subsequent three studies in the dissertation.

Study 3 contributes to the existing literature by developing and validating a comprehensive problem quality rating scale measuring five problem characteristics. Study 4 tests the reliability of the rating scale and suggests satisfactory inter-rater reliability of the rating scale measures and offers insights that students' perceptions of the problem quality may vary during the learning process. In addition, Study 4 shows that the measures using the problem quality rating scale are generally reliable and provide support for further validation in other PBL contexts. As far as we know, rating scales to measure problem quality have not been tested in different PBL contexts. Therefore this study adds to the current literature by taking the first step to developing a rating scale that is applicable across PBL contexts.

Study 5 demonstrates the utility of the developed rating scale in different contexts as well. The five characteristics measured by the rating scale are not only theoretically grounded (Dolmans et al., 1997; Hung, 2003), and empirically evidence-based (Des Marchais, 1999, Sockalingam & Schmidt, 2007) but they are also aligned with the constructivist principles (Savery & Duffy, 1995). This study also shows that the rating scale can be used to measure more aspects of students learning' using the rating scale, instead of being limited to knowledge tests or comparison of student-generated learning issues with faculty-intended learning issues. Furthermore, Study 5 adds evidence to the existing literature (Gijselaers & Schmidt, 1990; Soppe et al., 2005) that the extent of problem familiarity positively influences students' motivation and success in identifying the learning issues. It also reveals the unfamiliarity level of the problem may not be all that bad as it leads to thinking, questioning, and reasoning. Finally, it sheds light that more research work is needed to understand the interaction taking place in collaborative learning as a result of problem-solving. In sum, the consistency in the findings through the various studies using different approaches adds validity to the overall findings.

## IMPLICATIONS

### PROBLEM EVALUATION

With regards to problem evaluation, results from the first two studies suggest that the identified eleven characteristics contribute to the effectiveness of the problems. Of these the extent to which the problem leads to the intended learning issues is shown to be the most important to students and tutors. This adds further support to the existing understanding that the extent to which the problem leads to the intended learning issues is an important indicator of the problem effectiveness. In addition, the results add other characteristics that need to be considered in evaluating the effectiveness of problems. For instance, the function characteristics such as the extent to which the problem stimulates critical reasoning, promotes self-directed learning, stimulates elaboration, promotes teamwork, stimulates interest, and leads to the intended learning issues are likely to serve as appropriate indicators of problem effectiveness. One point to note is that these function characteristics represent the objectives of PBL (Savery and Duffy, 1995). Therefore, measuring these characteristics could be used to indicate to what extent the problem plays a role in the effectiveness of PBL.

Study 3 contributes by developing and validating a rating scale to measure five of these functional characteristics. Study 4 provides insights on the reliability of the rating scale to measure these problem characteristics. In addition, Study 4 shows that the problem quality rating scale can be used to measure students' conceptions of problem characteristics for problems across different PBL settings. To our knowledge, this is the first time that a rating scale for problem quality that has been tested in a particular PBL setting is applied in another PBL setting. This goes to support the generalizability of the findings. Nevertheless, further research using the problem quality rating scale in different PBL settings is needed to warrant the generalizability.

Study 5 shows the utility of the rating scale in measuring the influence of problem familiarity on students' learning. The comprehensive nature and psychometric characteristics of the rating scale supports use of the rating scale in future studies. Typically, students' performance on knowledge tests is used as measure of students' learning (e.g., Soppe et al., 2005). What this study adds is a possibility of using other measures of students learning which are in line with the principles of constructivist learning. In a way Study, 5 can be considered as a pilot study since the rating scale only tested on two problems. Success of this pilot study suggests that expansion of the study to a larger set of problems is feasible. Such studies will help shed more light on the impact of problem familiarity on students' learning. For instance, inclusion of more problems in the study would allow correlation of the students' learning with the tutors' assessment. Another implication of the results is that the rating scale can be used as formative evaluative tool to help provide feedback on individual problems. This will help problem designers in understanding which problems are effective and why.

## PROBLEM DESIGN

As for problem design, consistency of the eleven problem characteristics in the first two studies and comparability with literature (Des Marchais, 1999; Dolmans et al., 1997; Hung, 2003) suggests that these characteristics are likely to be useful in the design of problems. The classification of the characteristics as feature and function characteristics based on the assumption that the function characteristics can be manipulated by modifying the feature characteristics of the problems provide guidance on problem designs. For example, we can infer from the students' responses in Study 1 that the format of the problem can be modified to have an influence on students' interest in the problem. Context of the problem could be designed such that students are able to relate to some application or use in other modules/real-life context to have an impact on students' interest and learning. Appropriate key words and hints can be included in the problem to alter the clarity level of the problem so as to have an impact on the extent to which the problem leads to intended learning issues. The difficulty level of the problem can be adjusted to have an influence in the extent to which the problem is interesting, the extent to which the problem leads to learning issues, stimulates critical reasoning, and promotes self-directed learning.

Results from Studies 2 and 5 also support this assumption and reveal some useful information on problem design. For instance, problem format, which refers to how the problem is presented to students, such as the length of the problem, was found to influence the students' interest (Study 2). Specifically, longer problems were considered to be not highly motivating. Problems of relevant context (Study 2), and problems that matched students' prior knowledge in terms of content and context (Study 5) were found to be engaging. Problem clarity, which refers to how clearly the problem can be comprehended, was found to have an influence on the extent to which the problem leads to the intended learning issues. Problems that caused difficulty in comprehension were found to be ineffective in leading to the learning issues (Study 2). In line with this, Study 5 indicates that the problems with more keywords, clues, hints embedded in it guided students towards the learning goals. Even though problems that lacked in clarity (Study 2) and familiarity (Study 5) were found to be ineffective in guiding students towards the learning issues, these problems were found to trigger thinking, questioning, and reasoning. Problems designed to require multiple perspectives seemed to result in the same (Study 5).

Overall, the inference from Studies 1, 2, and 5 is that there exists a complex relationship between the feature and function characteristics. Hence, further research is needed to unravel this complex relationship to understand how to design effective problems. This dissertation makes initial attempts at unravelling some of this complex interrelationship by developing a tool in the form of a rating scale. However, it should be noted that findings in one PBL context may not be directly applicable in another context. Although Study 4 supports the reliability of the rating scale to measure the principal problem characteristics, further research is still needed to test the generalizability. Nevertheless, the value of these studies should be seen in the context of the suggested approach rather than just on the findings.

## LIMITATIONS

There are a number of limitations to the present study. First, the newly devised problem quality rating scale is limited to mostly measuring the function characteristics of problems. The rating scale does not measure the feature characteristics such as problem difficulty. Second, validation and reliability testing of the problem quality rating scale has been carried out with first-year students only. Studies by Sofie et al. (2007) suggest that students' conception of their learning changes with time. Hence, it will be relevant to investigate how students' perceptions about problem quality changes over time. Third, only limited numbers of problems have been evaluated with the rating scale. Hence, future research with more problems will be useful to gain more in depth understanding of problem design. Fourth, even though, tutor's perceptions were used to develop the rating scale, it has not been validated using tutors as respondents. Fifth, predictive validity of the rating scale remains to be tested.

## FUTURE WORK

To address these identified limitations, a number of future studies are proposed. First, the present rating scale can be expanded to measure the feature characteristics of problems. Alternatively, an experimental study approach can be used to investigate the influence of feature characteristic on students' learning, as in Study 5. Second, the utility of the rating scale needs to be assessed in varied conditions. For instance, students from different academic year and tutors can be involved in testing the rating scale. In addition, the rating scale can be used with several problems from different modules. Third, application of the rating scale to measure the problem quality formatively at problem level and summatively at modular level can be investigated. This proposal is currently being pursued at the Polytechnic in Singapore. Fourth, a study to assess the predictive validity of the rating scale can be carried out. Fifth, the rating scale can be validated and utilized in the other PBL context such as that in the Netherlands.

**REFERENCES**

Albanese, M. A., & Mitchell, S. (1993). Problem-based learning-a review of literature on its outcomes and implementation issues. *Academic Medicine, 68*, 52-81.

Alwis, W. A. M., & O'Grady, G. (2002). *One Day-One Problem at Republic Polytechnic*. Paper presented at the 4th Asia-Pacific Conference on PBL, 9-13 December, Hatyai, Thailand.

Anderson, J. R. (1990). *Meaning-based knowledge representations*. New York: Freeman.

Angeli, C. (2002). Teacher's practical theories for the design and implementation of problem-based learning. *Science Education International*, *13*(3), 9-15.

Araz, G., & Sungur, S. (2007). The interplay between cognitive and motivational variables in a problem-based learning environment. *Learning and Individual Differences, 17* (4), 291-297.

Arbuckle, J. L. (2003). Amos 5.0 Update to the Amos User's Guide: Marketing Dept., SPSS Inc.: Small Waters Corp.

Bandalos, D.L., & Finney, S.J. (2001). Item parcelling issues in structural equation modelling. In: G.A. Marcoulides & R.E. Schumaker (Eds.), *New Developments and Techniques in Structural Equation Modelling* (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum.

Barrows, H. S. (1986). A taxonomy of problem-based learning methods. *Medical Education, 20*, 481-486.

Barrows, H. S. (1990). Inquiry: The pedagogical importance of a skill central to clinical practice. *Medical Education, 24*, 3-5.

Barrows, H. S., Neufeld, V. R., Feightner, J. W., & Norman, G. R. (1978). *An analysis of the clinical methods of medical* *students and physicians*. Hamilton, Ontario, Canada McMaster University.

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning*. New York, NY: Springer Publishing.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.

Berlyne, D.E. (1978). Curiosity and learning. *Motivation and Emotion, 2*(2), 97-175.

Bloom B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain.* New York: David McKay Co Inc.

Blumberg, P., & Michael, J. A. (1992). Development of self-directed learning behaviours in a partially teacher-directed problem-based learning curriculum. *Teaching and Learning in Medicine,4*(1), 3-8.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Scienc*e, *16*, 153-184.

Boud, D., & Felleti, G. (Eds.). (1991). *The challenge of problem-based learning.* New York: St. Martin's Press.

Boud, D., & Feletti, G. (1997). *The challenge of problem-based learning* (2nd ed.). London: Kogan Page.

Brown, D.L. (1976). Faculty ratings and student grades: a university-wide multiple regression analysis. *Journal of Educational Psychology, 68*(5), 573-578.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*, 32-42.

Byrne, B. M. (2001). *Structural equation modeling with Amos. Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, *51*, 281-309.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education*. London: Routledge/Falmer.

Colliver, J. A. (2000). Effectiveness of problem-based learning curricula: Research and theory. *Academic Medicine, 75(*3), 259-266.

Custers, E. J., Boshuizen, H. P., & Schmidt, H. G. (1996). The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Memory and Cognition. 24,* 384–399.

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.

De Grave, W. S., Boshuizen, H. P. A., & Schmidt, H. G. (1996). Problem-based learning: cognitive and metacognitive processes during problem analysis. *Instructional Sciences, 24*, 321-341.

De Grave, W. S., De Volder, M. L., Gijselaers, W. H., & Damoiseaux, V. (1990). Peer teaching and problem-based learning: Tutor characteristics, tutor functioning, group functioning and student achievement. In Z. N. Nooman, H. G. Schmidt & E. S. Ezzat (Ed.). *Innovation in medical education: An evaluation of its present status, pp 123-135*. New York: Springer.

De Volder, M.L., Schmidt, H.G., Moust, J.H.C., & De Grave, W.S. (1986). Problem-based learning and intrinsic motivation. In

J.H.C. Van der Berchen, T.C.M. Bergen &E.E.I. De Bruyn (Eds.), *Achivement and Task Motivation* (pp.128-134). Berwyn, ILL: North America.

De Volder, M.L., Schmidt, H.G., Moust, J.H.C., & De Grave. W.S. (1986). Motivation and achievement in cooperative learning: the role of prior knowledge. In J.H.C. Van der Berchen, T.C.M. Bergen & E.E.I. De Bruyn (Eds.), *Achivement and Task Motivation* (pp. 123-127). Berwyn, ILL: North America.

Delisle, R. (1997). *How to use problem-based learning in the classroom.* Alexandria, Va.:Association for Supervision and Curriculum Development.

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

Dewey, J. (1916). *Democracy and education. An introduction to the philosophy of education* (1966 ed.), New York: Free Press.

Dochy, F.J. R.C., & Alexander, P.A. (1995). Mapping prior knowledge: a framework for discussion among researchers. *European Journal of Psychology of Education, 10*(3), 225-242.

Dochy, F., Segers, M., & Buehl, M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Education Research, 69*, 147-188.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction, 13*, 533-568.

Dolmans, D. H., De Grave, W., Wolfhagen, I. H. A. P, & Van der Vleuten, C. P.M. (2005). Problem-based learning: future challenges for educational practice and research. *Medical Education, 39*, 732-741.

Dolmans, D.H. J. M, Gijselaers, W. H., Moust, J. H. C., De Grave, W. S., Wolfhagen, I., & Van der Vleuten, C. P. M. (2002). Trends in research on the tutor in problem-based learning: Conclusions and implications for educational practice and research. *Medical Teacher, 24*, 173-180.

Dolmans, D.H. J. M., Gijselaers, W. H., Schmidt, H. G., & Van der Meer, S. B. (1993). Problem effectiveness in a course using problem-based learning. *Academic Medicine, 68*(3), 207-213.

Dolmans, D. H, J. M., & Schmidt, H. G. (1994). What drives the students in problem-based learning. *Medical Education, 28*, 372-380.

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*, 185-189.

Dolmans, D. H. J. M., & Wolfhagen, I. H. A. P. (2005). Complex interaction between tutor performance, tutorial group productivity and the effectiveness of PBL units as perceived by students. *Advances in Health Sciences Education, 10*, 253-261.

Dolmans, D. H. J. M., Wolfhagen, I. H, A. P., & Schmidt, H. G. (1996). Effects of tutor expertise on student performance in relation to prior knowledge and level of curricular structure. *Academic Medicine, 71*(9), 1008-1011.

Duch, B. J. (2001).Writing problems for deeper understanding. In B. J. Duch, S. E. Groh, &D. E. Allen (Eds.), *The power of problem-based learning: A practical "how to" for teaching* *undergraduate courses in any discipline* (pp 47–58). Sterling, Va.: Stylus.

Duch, B. J., Groh, S. E., & Allen, D. E. (2001). Why problem-based learning? A case study of institutional change in undergraduate education. In B. Duch, S. Groh, & D. Allen (Eds.), *The power of problem-based learning: A practical "how to" for teaching undergraduate courses in any discipline* (pp 47–58). Sterling, Va.: Stylus.

El Ansari, W., & Oskrochi R. (2004). What 'really' affects health professions students' satisfaction with their educational experience? Implications for practice and research. *Nurse Education Today, 24*(8), 644-655.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: an analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Engdahl, R.A., Keating J.R., & Perrachione J. (1993). Effects of grade feedback on student evaluation of instruction. *Journal of Management Education, 17(*2), 174-184.

Facione, P. A. (1990). *The Delphi report.* Millbrae, CA: The California Academic Press.

Facione, P., Facione, N., & Giancarlo, C. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic, 20*(1), 61-84.

Feldman, K.A. (1978). Course characteristics and college students' ratings of their teachers: what we know and what we don't. *Research in Higher Education, 9*, 199-242.

Gagne, R. M., Wager, W. W., Golas, K. C., & Keller, J. M. (2005*). Principles of instructional design.* Belmont, CA: Wadsworth Thomson Learning.

Gerzina, T. M, McLean, T, & Fairley, J. (2005). Dental clinical teaching: perceptions of students and teachers. *Journal of Dental Education, 69*(12), 1377-1384.

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.

Gonnella, J.S., Goran, M.J., Williamson, J.W., & Cotsonas, N.J., Jr. (1970). Evaluation of patient care. An approach. *The Journal of the American Medical Association, 214*(11), 2040-2043.

Greene, B.A., & Miller, R.B. (1996). Influences on achievement: goals, perceived ability and cognitive engagement. *Contemporary Educational Psychology, 21*, 181-192.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent systems. In R. Cudeck, S. d. Toit & D. Sörbom (Eds.), *Structural equation modeling: Present and future—A Festschrift in honor of Karl Jöreskog*. Lincolnwood, IL: Scientific Software International.

Harland, T. (2003). Vygotsky's zone of proximal development and problem-based learning: linking a theoretical concept with practice through action research. *Teaching in Higher Education, 8*, 263-272.

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.

Hogg, M A., & Cooper, J. (2003). *The SAGE handbook of Social Psychology*. SAGE, London.

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Hung, W. (2006). The 3C3R model: A conceptual framework for designing problems in PBL. *Interdisciplinary Journal of Problem-Based Learning, 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,* 1001-1007.

Jonas H. S., Etzel, S. I., & Barzansky B. (1989). Undergraduate medical education. *Journal of the American Medical Association, 262*, 1011-1019.

Jonassen, D. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development*, *45*(1), 65-94.

Jonassen, D. (1999). Designing constructivist learning environments. In C M. Reigeluth (Ed.). *Instructional-design theories and models* (pp 215-239). Mahwah, NJ: Lawrence Erlbaum Associates.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development, 48,* 63-85.

Jonassen, D & Hung, W. (2008). All problems are not equal: implications for problem-based learning, *Interdisciplinary Journal of Problem-Based Learning, 2*(2), 6-28.

Kamin, C. S., O'Sullivan, P. S., Younger, M., & Deterding, R. (2001). Measuring critical thinking in problem-based learning discourse. *Teaching and Learning in Medicine, 13*, 27-35.

Kim, S., Phillips, W. R., Pinsky, L., Brock, B., Phillips, K., & Keary, J. (2006). A conceptual framework for developing teaching cases: a review and synthesis of the 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.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction - Effects of direct instruction and discovery learning. *Psychological Science, 15*(10), 661-667.

Knowlton, D. S (2003). Preparing students for educated living: virtues of problem-based learning across the higher education curriculum. In D. S. Knowlton., & D. C. Sharp (Eds.). *Problem-based learning in the information age* (pp 5-12). San Francisco, CA; Jossey- Bass.

Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The Elementary School Journal, 94* (5), 483-497.

Krapp, A., Hidi, S., & Renninger, K. (1992). Interest, learning and development. In K. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3-25). Hillsdale, NJ: Erlbaum.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9,* 151-173.

Loyens, S. M. M., Rikers, R. M. J. P., & Schmidt, H. G. (2006). *Changes in students' conceptions of constructivist learning and the influence of the learning environment*. Doctoral dissertation. Rotterdam: Optima Grafische Communicatie.

Loyens, S. M. M., Rikers, R. M. J. P., & Schmidt, H. G. (2007). Students' conceptions of distinct constructivist assumptions. *European Journal of Psychology of Education, 12,* 179-199.

Loyens, S. M. M., Rikers, R. M. J. P., & Schmidt, H. G. (2007). The impact of students' conceptions of constructivist assumptions on academic achievement and dropout. *Studies in Higher Education, 32,* 581-602.

Majoor, G. D., Schmidt, H. G., Snellen-Balendong, H. A. M, Moust, J. H. C.,& Stalenhoef-Halling, B. (1990). Construction of problems for problem-based learning. In Z. Nooman, H. G. Schmidt & E. S. Ezzat (Eds.). *Innovation in Medical Education.*New York: Springer.

Mamede, S., Schmidt, H. G., & Norman, G. R. (2006). Innovations in problem-based learning: what can we learn from recent studies? *Advances in Health Sciences Education, 11,* 403-422.

Marin-Campos, Y., Mendoza-Morals, L., & Navarro-Hernandez, J. A. (2004). Students' assessment of problems in a problem-based pharmacology course. *Advances in Health Sciences Education, 9,* 299-307.

Maudsley, G. (1999). Do we all mean the same thing by “problem-based learning’? A review of the concepts and a formulation of the ground rules. *Academy of Medicine, 74*, 178-185.

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

Mayer, R. E. (1992). Cognition and instruction: On their historic meeting with educational psychology. *Journal of Educational Psychology, 84*, 405-412.

Mayer, R. E. (1999). Designing instruction for constructivist learning. In C. M. Reigeluth (Ed.). *Instructional-design theories and models* (pp 141-160). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

Neville, A. J. (1999). The problem-based learning tutor: Teacher? Facilitator? Evaluator? *Medical Teacher, 21*, 393-401.

Newman, M. (2003). *A pilot systematic review and meta-analysis on the effectiveness of problem-based learning.* Newcastle: Learning & Teaching Subject Network for Medicine, Dentistry and Veterinary Medicine.

Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis. *Medical Education, 15*(5), 315-322.

Norman, G. R., & Schmidt, H. G. (1992). The psychological basis of problem-based learning: a review of the evidence, *Academic Medicine, 67*, 557-565.

Norman, G. R., & Schmidt, H. G. (2000). Effectiveness of problem-based learning curricula:theory, practice and paper darts. *Medical Education*, *34*, 721-728.

Parsons. S., Croft. T., & Harrison, M. (2009). Does students' confidence in their ability in mathematics matter? *Teaching Mathematics and its Applications*, *28*, 53-68.

Reigulth, C. M., & Moore, C. (1999). Cognitive education and the cognitive domain. In C. M. Reigeluth (Ed.). *Instructional-design theories and models* (pp 51-68). Mahwah, NJ: Lawrence Erlbaum Associates.

Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 373-404). San Diego, CA: Academic Press.

Sanson-Fisher, R. W, & Lynagh, M. C. (2005). Problem-based learning: a dissemination success story? *The Medical Journal of Australia*, *183*, 258-260.

Savery, J. R. (2006). Overview of problem-based learning: definitions and distinctions. *The Interdisciplinary Journal of Problem-based Learning*, *1*(1), 9-20.

Savery, J. R., & Duffy, (1995). Problem-based learning: An instructional model and its constructivist framework. *Educational Technology*, *35*(5), 31-35.

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

Schmidt, H.G (1985). *Attributes of problems for problem based learning*. Unpublished manuscript. Maastricht, the Netherlands: University of Limburg.

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

Schmidt, H. G. (1994). Resolving inconsistencies in tutor expertise research: lack of structure causes students to seek tutor guidance. *Academic Medicine, 69*, 656-662.

Schmidt, H. G., & Boshuizen, H. P. A. (1990). *Effects of activation of prior knowledge on the recall of a clinincal case*. Paper presented at the Meeting of the American Educational Research Association.

Schmidt, H. G., De Volder, M. L., De Grave, W. S., Moust, J. H. C., & Patel, V. L. (1989). Explanatory models in the processing of science text : the role of prior knowledge activation through small-group discussion. *Journal of Educational Psychology, 81*, 610-619.

Schmidt, H. G., Dolmans, D. H. J. M., Gijselaers, W. H., & Des Marchais, J. E. (1995). Theory-guided design of a rating scale for course evaluation in problem-based curricula. *Teaching and Learning in Medicine, 7*(2), 82-91.

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

Schmidt, H. G., & Moust, J. H. C. (1995). What makes a tutor effective? A structural-equations modeling approach to learning in problem-based curricula. *Academic Medicine, 70*, 708-715.

Schmidt, H. G., & Moust, J. H. C. (2000). Factors affecting small-group tutorial learning: A review of research. In D. H. Evensen & C. E. Hmelo (Eds.), *Problem-based learning: A research perspective on learning interactions*. (pp 19-52). Mahwah, NJ: Lawrence Erlbaum.

Schmidt, H. G., & Moust, J. H. C. (2000). Towards a taxonomy of problems used in problem-based learning curricula. *Journal on Excellence in College Teaching, 11*(2), 67-72.

Schmidt, H. G., Norman, G., & Boshuizen, H. (1990). A cognitive perspective on medical expertise: theory and implication. *Academic Medicine, 65*(10), 611–621.

Schmidt, H. G., Vanderarend, A., Moust, J. H. C., Kokx, I., & Boon, L. (1993). Influence of tutors subject-matter expertise on student effort and achievement in problem-based learning. *Academic Medicine, 68*(10), 784-791.

Schmidt, H.G., Van der Molen, Te Winkel, & Wijnen (2009). Constructivist, problem-based learning does work: A meta-analysis of curricular comparisons involving a single medical school. *Journal of Educational Psychology.*

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

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475-522.

Slavin, R. E. (1996). Research on cooperative learning and achievement: what we know, what we need to know. *Contemporary Educational Psychology, 21*, 43-69.

Silen, C., & Uhlin, L. (2008). Self-directed learning - A learning issue for students and faculty! *Teaching in Higher Education, 13*(4), 461-475.

Sockalingam, N., & Schmidt, H. G. (2007). *Features of problems in problem-based learning: The students' perspective*. Paper presented at the HERDSA Annual Conference, Adelaide, Australia.

Sockalingam, N., & Schmidt, H. G. (2008). *Perspectives on the Problem Characteristics for Problem-Based Learning.* Paper presented at the Meeting of the American Educational Research Association.

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

Steinert, Y. (2004). Student perceptions of effective small-group teaching. *Medical Education, 38*, 286-293.

The Boyer Commission on Educating Undergraduates in the Research University. (1998). *Reinventing undergraduate education: A blue print for America's research universities.*New York: The Carnegie Foundation for the Advancement of Teaching.

Tiwari, A., Lai, P., So, M., & Yuen, K. (2006). A comparison of the effects of problem-based learning and lecturing on the development of students' critical thinking. *Medical Education, 40*(6), 547-554.

Toon, P. (1997). Educating doctors, to improve patient care - A choice between self directed learning and sitting in lecture struggling to stay awake. *British Medical Journal, 315*(7104), 326-326.

Torp, L., & Sage, S. (2002). *Problems as possibilities: Problem-based learning for K-16 education* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Uden, L., & Beaumont, C. (2006). *Technology and Problem-based Learning.* Hershey, PA: Idea Group Inc (IGI).

Van Berkel, H. J. M., & Dolmans, D. H. J. M. (2006). The influence of tutoring competencies on problems, group functioning and student achievement in problem-based learning. *Medical Education, 40*, 730-736.

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.

Van den Hurk, M. M., Wolfhagen, I. H. A. P., Dolmans, D. H. J. M., & Van der Vleuten, C. P. M. (1999). The impact of student-generated learning issues on individual study time and academic achievement. *Medical Education, 33*(11), 808-814.

Van der Vleuten, C. P. M. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education, 1*, 41– 67.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., Te Winkel, W. W. R., & Van der Hurk, M. M. (2006). Do student-defined learning issues increase quality and quantity of individual study? *Advances in Health Sciences Education, 11*, 337-347.

Vernon, D. T. A., & Blake, R. L. (1993). Does problem-based learning work? A meta-analysis of evaluation research. *Academic Medicine*, 68(7), 550-563.

Visschers-Pleijers, A. J. S. F., Dolmans, D. H. J. M., Wolfhagen, I. H. A. P., & Van der Vleuten. (2005). Student Perspectives on learning oriented interactions in the tutorial group. *Advances in Health Sciences Education, 10*, 23-25.

Voss, J. F., & Post, T. A. (1988). On the solving of ill-structured problems. In M.T.H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp 261-285). Hillsdale, NJ: Lawrence Erlbaum.

White, H. B. (1996). Dan tries problem-based learning: A case-study. In L. Richlin (Ed.), *To improve the academy* (Vol. 15,pp 75-19). Stillwater, OK: New Forums Press and the Professional and Organizational Network in Higher Education.

Williams, S. (1992). Putting case-based instruction into context: examples from legal and medical education. *Journal of the Learning Sciences, 2*(4), 367-427.

Williams, P., Iglesias, J., & Barak, M. (2008). Problem-based learning: application to technology education in three countries. *International Journal of Technology and Design Education, 18*(4), 319-335.

Woods, D. R. (1994). *Problem-based learning: how to gain the most from PBL*. Waterdown, Ontario: Donald R. Woods.

Yueng, E., Ay-Yueng, S., Chiu, T., Mok, N., & Lai, P. (2003). Problem design in problem-based learning: evaluating students' learning and self-directed learning practice. *Innovations in Education and Teaching International, 40* (3), 237-244.

Zanolli, M. B., Boshuizen, H.P., & De Grave, W. S. (2003). Students' and tutors' perceptions of problems in PBL tutorial groups at a Brazillian Medical School. *Educational Health (Abingdon), 15*(2), 189-201.

# APPENDICES

## APPENDIX A

*1. Problem should lead to the intended learning issues.* **“**Some problem trigger tends to give a lot of words while some give a little. However, what I want to see is key words in the problem statement. They do not have the need to be so obvious so that students will be able to search for resources immediately. However, key words which will give the students hints or even guide them to another major keyword and eventually allow them to find the key concept” (Participant B28).

*2. Problem should trigger interest.* “I would think that it is highly interactive and interesting when we are given problem statements that concern our everyday way of life” (Participant B1).

*3. Problem should be of suitable format such as length of text or use of visuals.* “My definition of good problem trigger is firstly, it has to be straight forward, NO NO NO to long winded ones, as the word 'trigger' tells all. It is the start of morning, a good problem can trigger off enthusiasm, if it is long winded, honestly, it can kill off the learning spirit” (Participant B5).

*4. Problem should stimulate critical thinking.* “My perception of a good problem trigger is one that actually gets you thinking. One that is 'not that obvious' but still not difficult to figure out what the problem is about” (Participant B 17).

*5. Problem should promote self-directed learning.* “Even though we complain that some of the problem triggers is difficult, I do think that is good, as difficult problem triggers activates our minds and we will not waste our time doing other stuff. Furthermore, when the problem is harder, we would always refer to it to make sure that we are not going off track. Easy ones might be neglected and at the end of the day, we may go too off track and learn things which are not related to the topic” (Participant B12).

*6. Problem should be of suitable difficulty.* "It would be good for the problem to be slightly difficult as what is the use of a problem trigger if it doesn't trigger the mind and make us think out of the box. I don't prefer problem triggers that are too easy and straight forward because it just seems too easy to be true and we might finish our task too fast. Thus, not making full use of the time given from the 2nd breakout till the 3rd meeting. Nevertheless, I do not prefer them to be too difficult because at times the topic that we need to touch on is quite a lot yet there is not much time to research and comprehend the findings before presenting" (Participant B5).
*7. Problem should be of suitable clarity.* "A good problem trigger must contain clue words of the topic being taught for the day. Even if it is without any help of the worksheet, at least we know what we had to learn" (Participant B11).
*8. Problem should enable application or use.* "The problem must be crafted in such a way that students would think out of the box in order to solve the trigger. If there are a lot of possible solutions, compared to always having one method in solving the trigger, the problem trigger would then be as challenging as it could be thought of. Having the knowledge of the lesson and solving is not enough. The students must be able to apply what they have learnt to their daily life. So that in future when students faced such problems in their workforce, they would be able to relate it to what they had learnt in school" (Participant B24).
*9. Problem should relate to prior knowledge.* "Problem should also relate to the real world, so that students have a stake in solving the problem. If at all possible, the problem should be placed in a context with which students are familiar" (Participant B21).
*10. Problem should stimulate elaboration.* "The problem trigger must be crafted in such a way that it is clear cut, easy to understand and contains keywords which are crucial to the day's problem. This would enable us to quickly start to research and brainstorm about

the various concepts and ideas of the day's lesson” (Participant B27).

11. *Problem should promote teamwork*. “Furthermore if it (problem) is difficult to a certain extent, it will enable us/me to think hard and at the same time have a better discussion within the team and class” (Participant B5).

Note: In the institute concerned, problems are also referred as problem statement and problem trigger.

## APPENDIX B

### 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.

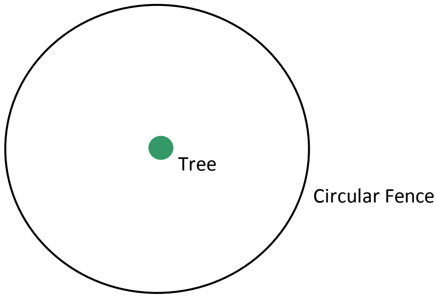
### Problem 04: Curious spots

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.

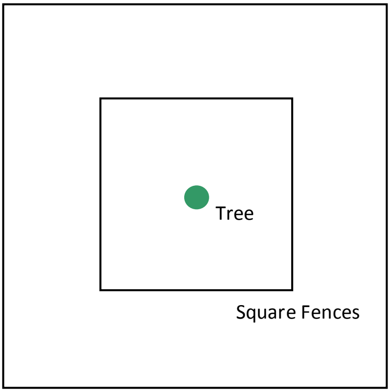
Over many generations, the Curions had tried in vain to look for a pattern in the way the spots appeared. They have since concluded that the spots appeared at random places along the fence determined by the will of the wise god.

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 the following diagram).




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.

**APPENDIX C**

*Detailed description of the five-factors and fourteen parcels*

| Parcels | Statement |
|---|---|
| Factor 1: The extent to which the problem leads to formulation of intended learning issues | |
| 1. Clarity of the problem | 1. I was clear about what the problem required my team and me to do*<br>2. The problem was clearly stated* |
| 2. Elements of clue or key words in problem | 3. The problem provided sufficient clues/hints*<br>4. The problem contained sufficient keywords* |
| 3. Structured approach to the problem | 5. I was able to identify the key learning issues from the problem*<br>6. I was able to come up with a satisfactory list of topics to explore on based on the problem*<br>7. I had a logical approach to the problem |
| Factor 2: The extent to which the problem is familiar to students | |
| 1. Familiarity with content | 1. I was familiar with the content of the problem even as I started to work on it*<br>2. I have personally experienced one or more situations described in the problem* |

| | |
|---|---|
| | 3. I could relate to the content of the problem based on my experiences* |
| 2. Relates to general knowledge | 4. The problem statement fits well with my prior knowledge*<br>5. The subject matter of the problem reflected current affairs/issues around the world* |
| 3. Relates to subject-domain knowledge | 6. I have done similar topic as in the problem before*<br>7. I had sufficient basic knowledge to identify suitable resources* |
| Factor 3: The extent to which the problem is interesting to students | |
| 1. Triggers personal interest at the start | 1. I was not interested to read the problem*<br>2. I was curious to find the answer* |
| 2. Engages in self-directed learning | 3. The problem stimulated me to find out more information on the topic*<br>4. The problem stimulated me to work hard during the breakouts |
| 3. Problem captivates attention | 5. The problem was engaging throughout the learning process<br>6. The problem captivated my attention throughout the day |
| Factor 4: The extent to which the problem promotes collaborative learning | |
| 1. Problem triggers brainstorming | 1. The problem triggered sufficient level of group discussion<br>2. We brainstormed over the problem on what we needed to find out |

| | |
|---|---|
| 2. Problem triggers team discussion | 3. Everyone in the team participated in the discussion<br>4. The problem stimulated us to discuss |
| 3. Problem encourages teamwork | 5. Team member's expertise in different subjects helped in solving the problem<br>6. Our team worked efficiently |
| Factor 5: The extent to which the problem stimulates critical reasoning | |
| 1. Problem stimulates thinking, questioning and reasoning | 1. The problem triggered lots of questions in my mind<br>2. I analyzed the information collected to respond to the problem<br>3. The problem stimulated me to think and reason statement |
| 2. Problem encourages multiple perspectives | 4. The problem had more than one right answer<br>5. There were many different viewpoints regarding the solution<br>6. Team members had diverse opinions on the problem |

* Items used in the shorter version of rating scale

## APPENDIX D

***Familiar problem used in the study – problem "Knowledge and morality"***

"Dad, what is business ethics?" the son asked.
"Well son, think that an old lady came to our store and mistakenly overpaid 500 dollars."
"Oh, yes, I can see. We tell she made a mistake or we keep quiet."
"No, son, business ethics is about whether we tell our business partners about the extra profit, or we keep quiet."
Examine the rights and wrongs of a choice based on what one knows.

***Unfamiliar problem used in the study – problem "Realism and anti-realism"***

Any suggestion to the effect that your hand phone is not real, or the person you talked to via the phone is imaginary, would sound lunatic. Nevertheless, if you ask how the hand phones work, you will hear things like photons leaving your phone and arriving at your phone – whether these photons are real things or some stuff imagined by scientists in order to explain what is going on, is an open question.
The debate of realism versus anti-realism in relation to human knowledge has raged for ages. Explore the issues addressed in this debate.

## CURRICULUM VITAE

Nachamma Sockalingam was born on May 24, 1974 in Madurai, India. She completed Bachelors of Science with Honours (Microbiology) in 1996 at the National University of Singapore. She did her Masters in Microbiology from 2000-2002 at the National University of Singapore. After successful completion of the project, she joined Johns Hopkins School of Medicine, Singapore, from 2002 to 2004 as a Research Assistant. In 2004, she joined Republic Polytechnic as an Academic Staff. From 2004 to 2008, she was attached to the School of Applied Science. During this time, in addition to facilitation, she supervised several student projects. She was then appointed to be the coordinator for students' final year projects in 2007. Meanwhile, her interest in educational research led her to explore how PBL problems could be designed for practical lessons. At this time, she had the opportunity to meet Professor Henk Schmidt, who encouraged her to investigate the characteristics of problems in problem-based learning as her PhD project. She is now a Senior Academic Staff at the Centre for Educational Development, Republic Polytechnic. Here, she has been involved in curriculum development, staff development, and educational research. In addition, she provides consultancy and conducts workshops on curriculum development and problem crafting. She is also the secretariat of the problem crafting certification panel in the Polytechnic. She continues to explore further on issues related to problem characteristics in PBL.

## PRESENTATIONS

Sockalingam, N., & Schmidt, H. G. (2009). *Development and validation of a questionnaire to assess problem familiarity, problem interestingness and extent to which problems lead to learning goals*. Paper presented at the Meeting of the American Educational Research Association, San Diego.

Sockalingam, N., & Schmidt, H. G. (2009). *Student and Tutor Perceptions of Effective Problems in PBL*. Paper *accepted* to be presented at the European Association of Research on Learning and Instruction Conference, Amsterdam.

Sockalingam, N., Rotgans, J. R., & Schmidt, H. G. (2009). *Development and Validation of a Questionnaire to Assess Problem Quality in PBL*. Paper *accepted* to be presented at the European Association of Research on Learning and Instruction Conference, Amsterdam.

Sockalingam, N., & Schmidt, H. G. (2009). *Eleven Characteristics of Effective Problems in PBL.* Paper presented to be presented at the Second International PBL Symposium, Singapore.

Sockalingam, N., & Schmidt, H. G. (2008). *Do students and tutors differ in their perceptions of problem effectiveness in Problem-Based Learning?* Paper presented at Teaching and Learning in Higher Education, Singapore.

Sockalingam, N., & Schmidt, H. G. (2008). *Perspectives on the Problem Characteristics for Problem-Based Learning.* Paper presented at the Meeting of the American Educational Research Association, New York.

Sockalingam, N., & Schmidt, H. G. (2007). *Features of problems for problem-based learning: The students' perspective.* Paper presented at the HERDSA Annual Conference. Adelaide, Australia

Sockalingam, N., & Schmidt, H. G. (2007). *Features of problems in problem-based learning: The students' perspective.* Paper presented at the International Problem-Based Learning Symposium, Singapore.

Sockalingam, N., Choo. S., Chua. A., Heng. J., & Chong, T. (2007). *The PBL Experience (Molecular Biology).* Workshop conducted at the International Problem-Based Learning Symposium, Singapore.