

**STUDENT SELF-
ASSESSMENT IN
HIGHER EDUCATION**

DUAN NING MAGDELEINE LEW



Copyright © 2009 D.N.M. Lew

All rights reserved. No part of this thesis may be reproduced or transmitted in any form, by any means, electronic or mechanical, without the prior written permission of the author, or where appropriate, of the publisher of the articles.

**Cover design and layout by TILT, Singapore
Printed by Ruby Printing Pte Ltd, Singapore**

Student Self-Assessment in Higher Education

Zelfbeoordeling van studenten in het Hoger Onderwijs

THESIS

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus
Prof.dr. S.W.J. Lamberts
and in accordance with the decision of the Doctorate Board

The public defence shall be held on
Tuesday 2nd of June 2009 at 11.00 hrs

By

Duan Ning Magdeleine Lew
born in Singapore



Doctoral Committee

Promotor: Prof.dr. H.G. Schmidt

Other members: Prof.dr. H.T. van der Molen
Prof.dr. R.M.J.P. Rikers
Prof.dr. C.P.M. van der Vleuten

CONTENTS

Foreword		9
Chapter 1:	Student Self-Assessment in Higher Education: An introduction	11
Chapter 2:	Reflection upon learning between theory and practice: A focus-group study of tutors' and students' perceptions	22
Chapter 3:	Measuring students' beliefs about self-assessment	47
Chapter 4:	Accuracy of students' self-assessment and their beliefs about its utility	70
Chapter 5:	Writing to learn and learning to write: Does reflection journal writing improve student learning?	100
Chapter 6:	Summary of research findings, implications and future research	122
References		138
Appendix A:	Self- and Peer assessment statements	146
Appendix B:	Questions for focus-groups with tutors	147
Appendix C:	Questions for focus-groups with students	148

Appendix D:	Self-assessment section of the Self-Assessment Questionnaire	149
Appendix E:	Grammar, Readability and Coherence tests	150
Curriculum Vitae, Publication and Presentations		154

For my dearest Mum

FOREWORD

I would like to express my heartfelt gratitude to the following individuals and institution who gave me the possibility to complete this thesis:

My promoter Prof. dr. Henk Schmidt: Dear Henk, I regard the last two and a half years as very educational and mind-stimulating. No words can express how thankful I am towards you for the insightful discussions, the brainstorming sessions for research ideas, stimulating suggestions, content-specific clarifications, helicopter view of things when I got lost in details, and encouragement you gave me in all the time of research for and writing of this thesis. You never fail to put a smile on my face after each research discussion. A lot of thanks for your faith and trust placed in me. I will never forget your kindness, which I will look back with pleasure.

Dr. W.A.M Alwis of Republic Polytechnic, Singapore: Dear Dr. Alwis, I am very grateful towards you for giving me the opportunity to pursue this Doctor of Philosophy degree. The last two and a half years working with you have been nothing but challenging (though sometimes nerve-wrecking!). You have not only been a great boss, but at the same time, an excellent mentor. Thank you for your faith in my abilities. I would not have been able to complete my research work if not for your continual support and encouragement.

My fellow colleagues from the Centre of Educational Development supported me in my research work. I want to thank them for all their insightful discussions, help, and encouragement. Especially I am obliged to Jerome Rotgans and Nicole Wang. I am grateful towards the assistance rendered by my colleagues from the Office of Information and Systems, for their help in administering the online questionnaires which were used as part of this research work, and for their assistance on data extraction.

My heartfelt thanks go to the Republic Polytechnic Singapore, where the research was carried out, and to its management staff who had made the data collection possible. Special thanks also go to all those staff and students who had contributed to this thesis.

I am also thankful towards my family, Mum (who is watching over me from Heaven), Dad, and younger brother Clarence, who have been great pillars of strength and emotional support during difficult times.

Especially, I would like to give my special thanks to Kenneth, whose patient love enabled me to complete this work.

Magdeleine
Singapore, March 2009

“The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvellous structure of reality. It is enough if one tries merely to comprehend a little of this mystery every day. Never lose a holy curiosity.”

Albert Einstein

Chapter 1- Student Self-Assessment in Higher Education: An introduction

This thesis addresses the theme of student self-assessment in higher education. Self-assessment is defined as the process by which students make judgments about their learning, particularly their learning outcomes (Boud & Falchikov, 1989; Eva et al., 2004). It functions to train students to make a better appraisal of aspects of their learning thus enabling them to take further steps to improve on their deficiencies (Thomas, 1999). Self-assessment is not only expected to encourage self-reflection or appraisal of their learning, it is also supposed to engage students actively in their learning. Falchikov (2005) contends that periodic self-assessment of learning processes and outcomes promote monitoring of learning progress, which in turn stimulates repair strategies that enable learners to further improve.

As the rapidity with which new knowledge is generated and disseminated becomes amplified, an increasing emphasis has been placed on the need for students to develop the skills necessary for effective learning and for successful functioning in professional practice. One such skill is the ability to self-assess (Eva et al., 2004). As such, the upsurge of interest in student self-assessment among researchers and educators in the past decade arises in part from the recognition of the positive role that self-assessment may play in student learning and in the development of professional competence. Several authors have argued that self-assessment is a critical tool for learning beyond university education, and that effective learning cannot take place without it. In his work, Boud (1989) emphasized that in the sphere of professional education, the need to monitor one's performance is one of the defining characteristics of professional work. Stefani (1994) further substantiates that the development of students' ability to assess and evaluate their work in ways applicable in their future profession is necessary for successful functioning in their careers.

1. Effects of self-assessment on student learning

Positive findings with regards to the use of classroom self-assessment have been reported in the literature. For instance, Dochy, Segers, and Sluijsmans (1999) analyzed 63 studies published between 1987-1998 on the use of self-, peer and co-assessment in higher education. Their review suggested that students who engaged in self-assessment demonstrated increased self-reflection and had better problem-solving skills. In another study, Orsmond, Merry, and Reiling (1997b) explored students' beliefs about the influence of self-assessment on their learning. Their findings revealed that students generally found the process of assessing their learning as challenging and beneficial, since it encouraged them to think critically and learn more which enabled them to work in a more structured manner (see also Segers and Dochy (2001)).

Self-assessment has been associated with moves towards developing greater student autonomy and responsibility in learning, particularly self-regulated learning (Paris & Cunningham, 1996; Paris & Paris, 2001). This form of learning emphasizes autonomy and control by the learner who monitors, directs, and regulates actions towards goals of information acquisition, expanding expertise and self-improvement. One of the critical self-regulatory skills that students need is the ability to self-assess. It is hoped that through self-assessment, students can internalize standards of professional expertise and reflect on their progress, enabling them to regulate their learning more effectively. In their work, Kraayenoord and Paris (1997) also emphasize that since self-assessment includes both self-reflection and evaluation of one's work, thus it can help to develop responsible and autonomous learners who are capable of regulating their learning.

Self-assessment has also been positively associated with enhancing metacognitive skills. In his work, Vockell (2004) describes metacognitive skills as the learners' automatic awareness of their knowledge and ability to understand, control and manipulate their cognitive processes. In reviewing the literature in the past century on teaching and learning, the American Psychological Association (1997) highlighted metacognition as one of the more important factors in

contributing towards effective learning. The review suggests that as students' metacognitive skills develop, so does their ability for self-reflection and self-regulation of learning, which in turn lead to improvements in academic performance. This is illustrated in the work by Lopez and Kossack (2007) who investigated the effects of continuous self-assessment on student course grades. They reported that the end-of-course correlations between students' self assessments and actual course grades were higher for the continuous self-assessment group, suggesting students were more realistically aware of their abilities when they periodically evaluate their understanding of course knowledge. In another study, Mok and her co-workers (2006) reported that the use of a metacognitive approach for self-assessment enhanced learners' awareness of their learning and processes of knowledge construction. The concept maps drawn by participants at the end of learning contained significantly more concepts and relationships than those drawn at the start of learning.

By contrast, some researchers are less optimistic about the use of self-assessment for improving student learning. For instance, Maguire, Evans, and Dyas (2001) demonstrated how first-year undergraduate students, when presented with self-assessment tasks, became 'strategic' in their approach to completing the tasks. Students were able to spot the possibility of achieving good results with minimal work and took advantage of that. Furthermore, they were sceptical about self-assessment and reflection, citing them as 'mechanical, meaningless tasks' which were non-beneficial to their learning. In another study, Maclellan (2001) compared teachers' and students' perceptions of self-assessment. Her study revealed that although teachers reported that they understood the purpose of self-assessment, it was infrequently used and, if so, exclusively at the end of a module. Students reported that they did not exploit self-assessment to improve their learning and furthermore, appeared to have an underdeveloped conception of what self-assessment was.

2. Student self-assessment accuracy

Besides studies which examined the effects of self-assessment on student learning as those reviewed above, there is also a body of literature reporting empirical studies that compare student-provided marks with those of teachers. In light of this type of self-assessment, research usually looks into the validity of the grades, by comparing the accuracy of the grade given by the learner with that awarded by teachers or peers (Boud & Falchikov, 1989; Falchikov & Boud, 1989). These studies have consistently showed that despite the accepted theoretical value of self-assessment on learning, the accuracy of student self-assessment is poor. For instance, Cassidy (2007) examined the self-assessment ability of first-year university students from a department of health sciences. Students were asked to provide marks for their work which were then compared with tutors' actual marks. A fairly low correlation of .25 was obtained between student-estimated and tutor marks. A more detailed analysis of students' estimates revealed that the majority of the students (56%) underestimated their assignment marks compared with 40% of students who overestimated their marks. Based on these findings, Cassidy concluded that the majority of students exhibited a good level of self-assessment skills, with a quarter of them failing to demonstrate such skills. He goes on to contend that such an observation may be testament to a changing trend in higher education which exposes students to teaching practices which involve students in making judgments about their learning and performance, thus assist in the development of self-assessment competency.

Some researchers have also investigated the stability of student self-assessment ability over time. In the study by Fitzgerald, White, and Gruppen (2003), they reported moderately high correlations ranging from .46 to .69 between students' estimated self assessments on knowledge examinations and their actual examination scores in the first two years of medical school. Lower correlations ranging from .12 to .42 were obtained for students' estimated self assessments on clinical examinations and their actual performance in their third year. The findings when taken together suggest that medical student self-

assessment accuracy is reasonably stable when compared to the stability of actual performance. In another study, Eva et al. (2004) asked all students enrolled in a medical program to rate the strength of their understanding of broad medical knowledge relative to their peers. The correlations between students' self assessments as compared to their actual performance scores were low, ranging from $-.24$ to $.13$. Furthermore, the findings the study suggest that students' ability to assess their learning deficiencies did not improve with experience with the self-assessment activity, despite receiving substantial feedback from their teachers regarding previous test performances.

Besides self-assessment, peer assessment has also received much attention in higher education. Boud (1986) suggests that the contribution of other students can be a very useful input into the self-assessment process. Learners have an opportunity to observe their peers throughout the learning process and often have a more detailed knowledge of the work of others than do their teachers. Boud further contends that peer assessment should not be considered only as a grading procedure; it should be seen as part of the self-assessment process and which "serves to inform self-assessment" (p. 22). Nicol and Milligan (2006) substantiate further that through reflecting and evaluating the performance of their peers, learners are expected to develop objectivity in relation to standards which can be then transferred to their own work. Other beneficial effects of peer assessment on student learning reported in the literature include: improved understanding of subject matter, enhanced metacognitive understanding of the learning process and encouraged the development of skills of reflection and critical reasoning skills (McDowell, 1995; Race, 1998; Searby & Ewers, 1997).

The literature also reports of empirical studies which compared peer awarded marks with student-estimated or tutor marks. For example, a qualitative review by Topping (1998) that focused primarily on the mechanisms and benefits of peer assessment located 25 studies, that compared teacher and peer marks, and 8 studies, which compared student and peer marks. The majority of the studies (18) demonstrate that peer assessments were of adequate reliability when compared to student and teacher marks in a variety

of applications. Similar findings were mirrored in the meta-analysis conducted by Falchikov and Goldfinch (2000).

3. Reflection journal writing as a form of self-assessment

Reflection journal writing also provides many opportunities for students to engage in self-assessment. Students reflect on their learning process and achievements and document these reflections, understand their learning progress through journal keeping, and share their personal responses to collaborative learning and school work with their teachers (Chirema, 2007; O'Connell & Dymont, 2006). Beneficial effects of journal writing on student learning reported in the literature include: encouraging the development of self-reflection and metacognition, and to improve writing skills (Moon, 1999a). Nonetheless, there are also problems associated with the use of journals as learning tools in higher education, namely, the use of journals as tools to criticize fellow students and which were expressly written for what the teacher wished to see (Kerka, 1996).

The studies reviewed above provide a mixed picture regarding the utility of self-assessment tools in higher education, since researchers reported beneficial effects of these tools on student learning, while others were less optimistic. Empirical studies which examined student self-assessment accuracy often involved only a limited number of participants and compared student self-judgments with single teacher and peer judgments. Such judgments by teachers and peers usually take place only towards the end of a task or course and are thus limited by recall (Falchikov, 2005). The research presented in this thesis was thus aimed at exploring whether the seemingly mixed results of the studies reviewed, as well as the reliability and validity issues associated with studies on student self-assessment accuracy, can be resolved in a curriculum where self-assessment is a continuous (i.e. daily) activity.

4. The Republic Polytechnic problem-based curriculum

The polytechnic at which the research was carried out organizes its curriculum according to principles of problem-based learning. Here,

students work collaboratively in teams of four to five, with learning centred on problems relevant to their domains of study. Students work each day on one problem during a five-day work week. The problem is initially discussed in the morning, followed by ample study. At the end of the day, information gathered is shared and elaborated upon. No didactic teaching takes place or is there any form of direct instruction. One tutor supervises the student teams in a larger classroom (Alwis, 2007).

Assessment at the polytechnic involves students being graded daily, and they having to take knowledge acquisition tests. The daily assessment approach consists of four, independent elements: (1) a self-assessment, (2) a peer assessment (3) a reflection journal, and (4) a judgment by the tutor on how well students have performed during the day. The self-assessment rating scale consists of 8 items inquiring about the quality of students' performance within their team, such as the level of cooperativeness and contribution of ideas. A Cronbach's alpha value of .90 gives evidence for the high internal consistency reliability of the self-assessment instrument. The peer assessment rating scale consists of 4 items inquiring about the cooperativeness and quality of contributions of peers within the team. The peer assessment instrument has high internal consistency reliability, given its Cronbach's alpha of .93. In examining the inter-rater agreement by correlating the scores awarded to students by different peers, we computed the intraclass correlations based on students' peer assessment scores. Intraclass correlations of .97 and .95 for the first and the second semester respectively were obtained. The values of Cronbach's alpha were computed based on student responses on the items of the self- and peer assessment instruments in semester one of the 2007-2008 academic year. Students are asked to respond to these items on a Likert five-point scale ranging from "strongly agree", "disagree", "neutral" and "agree" to "strongly agree". The items for the self- and peer assessment are contained in Appendix A. On a particular day, each student assesses and is in turn assessed by his peers within the team.

Students' reflection journals form a part of the daily assessment approach. The reflection journal is a short essay created by the student that is "personal" and records his or her daily reflections of

daily learning in respond to a reflection journal question provided by the tutor. Each student is required to submit his or her reflection journal by the end of the day. Tutor-asked journal questions mainly required students to be reflective about their learning and development. Some examples of journal questions include "What are some of the strengths that I demonstrated today?", "What insights did I gain today?", "What strategies have I used to help me in my learning", "What prior knowledge did I apply to help me understand the problem better?" and so on. Students respond to a different reflection journal question each day. The didactic purpose of writing the reflection journal is in line with the literature reviewed above, to encourage and record self-reflection about the process of learning.

The tutor judgment consists mainly of tutors' observations of students' processes of daily learning. The observations by the tutors include students' self-directedness, level of participation inclusive of teamwork; students' ability to reason, justify and defend opinions and ideas formulated in respond to problems, as well as their problem solving skills. Tutors will then award grades ranging from "A" to "F", which are derived based on what they observe and the impression they have on each student during the duration of time they had with him/her. Tutors also take into consideration students' individual reflection journals (short essays which document students' reflections on daily learning) and their self and peer assessments when awarding grades. Furthermore, tutors will provide feedback to students on their learning outcomes and processes of daily learning. The generalizability of judgments made by different tutors is high, with an average generalizability coefficient of .84 (Chua & Schmidt, 2007).

Students also need to take four knowledge acquisition tests per module, which are taken at different points during the semester. The duration of each test is 30 minutes and it consists of answering at least three structured questions. The tests are conducted in a supervised environment, similar to an end-of-course examination. Students are tested on their ability to understand and apply what they have learnt.

To that end, four studies have been conducted that constitute attempts to address the following research questions: (1) How do

students and tutors differ in their views about the purposes and utilities of the different self-assessment tools?; (2) What are students' beliefs about the utility of self-assessment?; (3) How accurate are students' self assessments as compared to peer and tutor assessments?; and (4) Does reflection journal writing improve student learning?

5. Research questions

Research question 1: How do students and tutors differ in their views about the purposes and utilities of self-assessment tools?

Response to this question might be considered as the starting point for research on the theme of student self-assessment in higher education. Given the impact of any form of assessment on student learning, many researchers (e.g. Langer, 2002; Segers & Dochy, 2001) have emphasized the importance of striving to obtain a match between the didactic functions of the different self-assessment tools and how students make use of them in reality. In order to fully exploit the effects of self-assessment on student learning, one must first understand how students perceive these assessment tools. Langer (2002) contends that the way in which a student perceives assessment will determine the way he or she responds to it. To add on, the learner's experience of assessment will determine the manner in which he or she tackles his or her learning.

Aimed at answering my first research question, I conducted a focus-group study with tutors and students to elicit their views about the purposes and utilities of self-assessment tools, namely, self-assessment and reflection journal. Tutors and students also shared their views about the purposes and utilities of two other assessment tools: peer assessment and tutor judgment. **Chapter 2** reports the findings of this study.

Question 2: What are students' beliefs about the utility of self-assessment?

The studies reviewed were largely based on *ad hoc* questioning of student and teachers about the utility of self-assessment. A calibrated instrument to measure students' beliefs about the effects of self-assessment tools on their learning is lacking. The absence in the literature of a validated instrument to measure students' beliefs about the utility of self-assessment practices on their learning was the motivation behind the second study reported in **Chapter 3**. The second research question was addressed by conducting a validation study on a questionnaire containing statements inquiring students about their beliefs about the utility of self-assessment and reflection journals on their learning. The hypothesized questionnaire model contained seven factors based on belief categories derived from past research studies as well as findings from the first study on the use of these self-assessment tools in higher education. In seeking evidence for multigroup invariance, the hypothesized questionnaire model was also tested on another independent student sample of the same population. Furthermore, the measurement stability of students' beliefs over time (i.e. test-retest reliability) was also investigated.

Research question 3: How accurate are students' self assessments as compared to peer and tutor assessments?

How accurate is self-assessment of the process of learning? Do students judged to be more academically capable have better developed self-assessment skills? Does the accuracy of self-assessment improve with time, experience and continuous feedback from teachers and peers? Providing responses to these questions would enrich current understanding of student self-assessment accuracy. The relationships between students' experience with self-assessing their learning, their intellectual capacity and their ability to assess themselves accurately would be further clarified. To evaluate student self-assessment accuracy, their self assessment scores were compared with multiple judgments of teachers and peers. The accuracy of students' self assessments was studied throughout a

semester in which the students involved made approximately eighty self assessments each. This study was conducted to gather evidence as to whether students can learn to self-assess given that they repeatedly have to evaluate themselves and receive continuous feedback from their teachers and peers.

I also examined whether self assessments were more accurate if students believed that this activity contributed to improving their learning. It was conjectured that a relationship exists between students' beliefs about the effects of self-assessment on their learning and their ability to self-assess accurately. It was hypothesized that students who believe that their learning improves through self-assessment are expected to achieve better performance grades. **Chapter 4** reports the findings.

Research question 4: Does reflection journal writing improve student learning?

Is there any evidence of learning in student reflection journals? Does students' awareness of their learning differ significantly with continuous engagement in journal writing and receiving feedback from teachers? To test these questions, student submitted reflection journals at the beginning and towards the end of an academic year were analyzed by means of text analysis software. I was also interested to investigate whether journal writing improve students' writing skills. To that end, spelling, grammar, readability and coherence tests were performed on student journals in seeking evidence of writing skills. **Chapter 5** reports the findings of this study.

Finally, **Chapter 6** summarizes the findings of this research and discusses their significance. In addition, the chapter suggests issues to be further explored for future research. The chapter concludes with implications of the findings for educational practice.

Chapter 2- Reflection upon learning between theory and practice: A focus-group study of tutors' and students' perceptions

ABSTRACT

This paper presents findings from a focus-group study that assembled tutors' and students' perceptions of the use of reflection upon learning as a tool for assessment at an institution of post-secondary education using problem-based learning. Assessment is conducted daily and consists of: (1) a self-assessment activity, (2) a peer assessment activity, (3) writing a reflection journal, and (4) a tutor judgment (i.e. the "daily grade"). Qualitative analyses of collected data from focus-group interviews with tutors ($n_1=7$) and students ($n_2 =15$) revealed that they understand the purposes of the various assessments but their perceptions of their utility differ. Tutors generally believe that a multifaceted approach provides a rich understanding of how well students are learning and self-reflection may help students to become better learners. In contrast, students cannot see the various assessments as valuable in their own right (as was the purpose). They prefer to reflect on content (rather than process) of their learning, and mainly worry about their daily grades. Students also believed that their self assessments were used by their tutors (to some extent) in arriving them. Under such conditions, self-assessment takes the risk of becoming a tool for impression formation.

Keywords: *Daily assessment; problem-based learning; self-assessment; peer assessment; reflection journal*

1. INTRODUCTION

New modes of assessment have enriched the 'conventional' evaluation setting, formerly characterized by the multiple-choice and essay tests, and by end-of-course examinations (Sambell, McDowell, & Brown 1997). A problem with conventional assessment practices is that they tend to be oriented towards tests of factual knowledge. Hence, they are considered limited by many researchers for their inability to adequately test for and measure higher order skills (Falchikov, 2005; Scouler, 1998; Segers & Dochy 2001). This pertains in particular to (i) cognitive skills: such as reflection, problem solving and self-directedness; and (ii) interpersonal skills: such as team work, leadership, helping peers, and communication skills. These competencies, however important for success in the workplace they may be, are seldom assessed. This is particularly problematic while implementing problem-based learning as an instructional and learning method, which focuses explicitly on the development of such cognitive and interpersonal skills (Schmidt, Vermeulen, & van der Molen, 2006). Problem-based learning tend to be characterized by students working collaboratively in small groups, with learning centred on problems relevant to the students' domain of study and much time spent on self-directed learning. The pedagogical appeal of problem-based learning is its perceived capacity to foster, through these learning processes, enhanced clinical reasoning skills, and the development of both an adaptable knowledge base to use in professional settings and the skills in self-directed learning necessary to become lifelong learners in that profession (Kelson & Distlehorst, 2000). Hence, assessing higher order competencies in such an approach seems to be called for.

The failure of conventional assessment to measure higher order competencies have resulted in the introduction of alternative assessment methods, such as self- and peer assessment tools, and reflection journals (Moon, 1999a; Segers & Dochy, 2001). These methods seek not only to encourage reflection, the ability to evaluate oneself and one's peers, but also serve to actively engage the students in their own learning process. However, the introduction of alternative assessment tools may solve some of the old problems;

they also may create new problems. For instance, Struyven, Dochy, and Janssens (2002) reported concerns of students over the fairness and authenticity of such assessment. Maclellan (2001) has argued that students do not exploit assessment to improve their learning and appeared to have only a surface level of understanding of the functions of assessment. The purpose of the study to be presented here was to find out to what extent a curriculum-wide introduction of assessment aimed at self-reflection causes such new problems to arise, as seen through the eyes of tutors and students. Therefore, we will briefly describe and discuss three of such assessment tools: self-assessment, peer assessment, and reflection journals.

Self-assessment. Self-assessment is defined as the process by which students make judgments about their own learning, particularly their learning outcomes (Boud, 1986; Eva et al., 2004). It functions to train students to make a better appraisal of aspects of their own learning and enabling them to take further steps to improve on their learning deficiencies (Thomas, 1999). For instance, Orsmond, Merry and Reiling, (1997b) divided 105 students in a course on physiology into small groups of 2-3 students to complete a poster assignment. Students were asked to complete a 13-item evaluation questionnaire at the end of their poster presentation. The questionnaire elicited students' responses about the influence of self-assessment on their learning, and their experience with the self-assessment exercise. The outcomes of data analyses demonstrated that the majority of the students reported that self-assessment made them (i) think and learn more, (ii) more critical of their own work, and (iii) enabled them to work in a more structured way. Students also perceived the self-assessment activity as being challenging, helpful and beneficial to improving learning.

In their study, Struyven et al. (2002) reviewed 36 empirical studies which explored students' views about alternative forms of assessment, including self- and peer assessment, and reflection journals, as compared to the more common multiple-choice and essay examinations. In general, alternative assessment was perceived by students to enable, rather than to 'pollute', the quality of learning achieved. Many students emphasized that for alternative assessment, they were channelling their efforts into trying to understand, rather

than simply memorize or routinely document, the material being studied. The findings of the study by Struyven and co-workers also indicated that students believed that alternative assessment was fairer, since it appeared to measure qualities, skills and competencies which would be valuable in contexts other than the immediate context of assessment.

In another study, Mires and others (2001) reported on a pilot study undertaken to evaluate the feasibility and reliability of undergraduate medical students' self-marking of degree written examinations, and to survey student opinion regarding the process. Their study revealed that students were concerned over the reliability of self and peer assessment, even though they valued these activities. They also had mixed feelings about being capable of assessing their peers fairly, in spite of the interrelated correlations between peer and tutor scores. Nonetheless, students perceived more disadvantages (e.g., being more stressful, uncertainty about capability, not knowing how to mark etc.) than advantages (e.g. seeing mistakes, feedback, and as a learning opportunity) in the self-marking exercise (see also Segers and Dochy (2001)).

Peer assessment. Peer assessment engages students in the process of judgment making about the performance of their peers (Sluijsmans et al., 2001). It requires students to provide feedback or grades (or both) to their peers on a product, process, or performance, based on the criteria of excellence for that product or event which students may or may not be involved concomitantly in determining the criteria (Falchikov, 2005). Implicit in the design of peer assessment is the assumption that students will be accurate and fair when assessing their peers. This, it is claimed, encourages students to develop high levels of responsibilities and a sense of ownership for their peers' learning (Dochy, Segers, & Sluijsmans, 1999). Nicol and Milligan (2006) further substantiate that as peer assessment engages students in the process of reflecting on and evaluating the performance of others, they can develop objectivity in relation to standards which can be then be transferred to their own work.

Beneficial effects of peer assessment on student learning have repeatedly been reported in the literature. For instance, Ballantyne,

Hughes and Mylonas (2002) reported that peer assessment enhanced the metacognition of learners and improved understanding of subject matter. Peer assessment was regarded as an awareness raising exercise which enabled students to consider their own work more closely, highlighted what they needed to know in the subject, helped them make a realistic assessment of their own abilities, and provided them with skills that would be valuable in the future. Furthermore, the peer assessment process also encouraged students to compare and reflect on their own work, which in turn is an important element of self-directed learning. More recently, Papinczak, Young and Groves (2007) conducted a qualitative study of first-year medical students' attitudes to, and perceptions of, peer assessment. Their study revealed that students felt a sense of increased responsibility towards their peers' learning. Students also reported that feedback from peers also assisted them in identifying deficiencies in their understanding and skills that were not readily apparent, thereby enabling them to take steps to further improve (see also Nicol and Milligan (2006)).

Although the studies mentioned thus far seem to suggest support for peer assessment, there are, however, several problems and limitations that have repeatedly been associated with the process of assessing others. For example, evaluation of student progress in a student-centred curriculum like problem-based learning, however, has remained a challenge (Eva et al., 2004). This is because teachers often develop assessment procedures that test content knowledge (e.g. end-of-course examinations), rather than on areas like problem solving, and skills as a group member. Eva and others (2004) suggest that the assessment of student skills, processes, and attitudes in problem-based learning schools will take place most appropriately within the tutorial setting. They go on to substantiate that since there are many opportunities to assess areas of student proficiency such as communication skills, teamwork, and respect for others (which are not readily evaluated by other forms of content knowledge-based tests) in the tutorial setting, problem-based learning educators are strongly encouraged to adopt tutorial-based peer assessment in their classrooms.

Another ramification arising from the findings of research on peer assessment is that students often lack confidence in their own and their peers' abilities as assessors. Students frequently report feeling "uncomfortable" in carrying out peer assessment, often because they feel unqualified to make these judgments (Mires et al., 2001). In another study, Sluijsmans et al. (2001) confirmed the existence of bias in peer marking due to interpersonal relationships between students. Furthermore, Papinczak and others (2007) highlighted that students were overt in their scepticism towards the peer assessment exercise. Students' verbatim responses such as "not taken too seriously" and "not much thought goes into the marking" reflected their casual attitudes towards the peer assessment process.

Most of the studies found in the literature on peer assessment focuses on the evaluation of individual contributions to group assignments or the validity and reliability of peer assessment. Although student perceptions of the peer assessment experience have been studied extensively in higher education, few studies have been concerned with evaluating students' views in a problem-based learning tutorial setting (Eva et al., 2004). To add on, many of the studies reportedly used peer assessment for summative purposes to judge the product of collaborative work (e.g., a poster or report), and is mainly administered towards the end of a predefined period for judging the quality of peers' works (Ballantyne et al., 2002).

Reflection journals. Reflection journals are sometimes called learning journals (e.g. Creme, 2005; Langer, 2002) or reflective journals (e.g. Thorpe, 2004). Education practitioners use reflection journals as tools to encourage and record reflection on learning (Meyers & Thomas, 1993; Moon, 1999b; Wong, Kember, Chung, & Yan, 1995). Although they are used in a variety of courses and for different purposes, reflection journals are essentially written records that students create as they think about various concepts learnt, critical incidents involving their learning, or interactions between students and teachers, over a period of time for the purposes of gaining insights into their own learning (Grant, Kinnersley, Metcalf, Pill, & Houston, 2006). For instance, Wong et al.,(1995) in a course on nursing, instructed students who were registered nurses, to submit learning journals at the end of the course. Students were asked to

reflect about how to apply theory into practice at their workplace. Content analyses of 45 learning journals suggested that student writing can be used as evidence to indicate the presence or absence of critical thinking and reflection.

In another study, Langer (2002) investigated how adult learners in a computer technology class responded to the use of learning journals. Students had to submit weekly journals that asked them to be reflective about new career opportunities and how to apply technology to the workplace. A qualitative review of 300 journals submitted by 20 students and that of interviews with ten student volunteers who were solicited six months after completing the course was performed. Langer's study revealed that adult learners used journals as study tools to gauge understanding of subject matter and in preparing for examinations. More recently, O'Connell and Dymont (2006) explored the experiences of 8 university faculty members who used learning journals in courses on outdoor education. A 32-item questionnaire that questioned faculty on their and their students' current practices of journal writing, and a focus-group discussion that asked faculty on their perceptions of journal writing were used for this study. The findings revealed that faculty members acknowledged the use of learning journals as a tool to encourage reflection and learning in students.

Grant et al. (2006) also investigated the effects of journal writing on students' learning. Twenty third-year undergraduate medical students who volunteered to participate in the study were asked to keep a learning journal for two semesters, and attended fortnightly, facilitated tutorial groups where they discussed their journal entries. Interviews with 19 students were conducted, where they were asked about their experiences with the tutorial groups, journal writing and how they felt they benefited from the study. Qualitative analyses of the collected data revealed that through journal keeping, students became better aware of their individual learning style, improved in the way they integrate what they learn from different sources, and felt a sense of personal achievement in what they knew.

The introduction of new assessment instruments assumes that teachers and students understand what the *raison d'être* of these instruments are. Do teachers and students understand the purposes

and functions of these new approaches? If yes, what then is their level of understanding? For instance, Maclellan (2001) administered a 40-item questionnaire which included items on the purpose of assessment, the nature and demand level of the tasks which were assessed, the timing of assessment and the procedures for marking and reporting, to 80 tutors and 130 undergraduate students. The data showed that students did not exploit assessment to improve their learning and appeared to have only surface understanding about the functions of assessment. Although tutors maintained that they understood the purposes of assessment, they reported that assessment never took place at the start of a module, students were not allowed to be assessed even though they felt ready for assessment and self- and peer assessment were infrequently used. Staff also reported to have placed heavy emphasis on the grading and ranking of students instead of focusing on motivating students and providing feedback on their learning.

An implication which arises from the study by Langer (2002) on the use of learning journals in continuing higher education is that student perception and scepticism of the assignment can affect the objective of developing critical thinking. Boud and Falchikov (1989) in their review of previous research on peer assessment from 1932 to 1988 reported that although students generally displayed a liking for peer assessment, they are less appreciative of criticism from their peers, felt uncomfortable in making negative judgments, and were unconfident and frustrated when rating the works of their peers due to insufficient practice or guidelines on how to go about peer assessment (Segers & Dochy, 2001; Sluijsmans et al., 2001). Students also expressed concern with the reliability of peer ratings as it was difficult to ameliorate rating errors like friendship marking, which results in over-marking, collusive marking, resulting in a lack of differentiation within groups, decibel marking, where individuals who dominate the groups get the highest marks; and parasite marking, where students fail to contribute but benefit from group marks (Pond, Ul-Haq, & Wade, 1995).

The studies reviewed above provide a mixed picture of the utility of self-assessment in higher education as perceived by teachers and students. Some authors report beneficial effects of self- and peer

assessment, and reflection journals (e.g. Moon, 1999a; O'Connell & Dymont, 2006; Orsmond, Merry, & Reiling, 1997a), while others are less optimistic (e.g. Boud & Falchikov, 1989; Maclellan, 2001; Ward, Gruppen, & Regehr, 2002). There may be several reasons for this undesirable state of affairs. The first is that at least in some studies students only had a superficial understanding of the purpose of self-assessment practices (Maclellan, 2001). If students do not understand *why* they have to assess their own learning, one cannot expect such approach to be attaining its goals, i.e., helping students to improve on their learning. The second reason may be that in quite some studies, self-assessment took place only at the end of a course. Of course, in such circumstances assessment cannot contribute to learning, simply because the learning is already over. A third reason could be that measures of self-assessment were, directly or indirectly, used for summative decisions about students (Boud & Falchikov, 1989; Segers & Dochy, 2001; Ward et al., 2002). *Summative* assessment, as opposed to formative assessment, aims at deciding whether students have passed an examination, can continue with their studies, or have reached a minimum level of proficiency. *Formative* assessment, by contrast, aims at providing students with feedback helping them to improve on learning. One cannot exclude the possibility that summative use of self-assessment may change the purpose of the whole exercise for students.

The purpose of the present study was to find out whether the possible shortcomings of the approaches described above, and the seemingly mixed results of the studies reviewed, can be resolved in a curriculum in which self-assessment is a continuous (i.e., daily) activity, and assessment is largely formative. Would teachers and students, under such conditions, understand the purpose of self-assessment, and would students feel supported in their learning? To that end, we studied the functions of assessment in an institution of post-secondary education that uses problem-based learning and in which students reflect on their learning on a daily basis. In the particular curriculum, students evaluate their own performances, are judged by their peers, write learning journals, and receive feedback from their tutors. The purpose of this study was to describe daily assessment methods as these were experienced by tutors and

students. To that end, focus-group interviews with students and tutors were conducted and the data collected were qualitatively analyzed and reported.

2. METHOD

2.1 Subjects

Subjects were 7 tutors and 15 students, who were invited to participate in this study in August, 2006. The tutors were selected using a purposeful sampling approach, based on the following selection criteria: (1) they had taken classes for at least two semesters; (2) they were familiar with the daily assessment procedure; and (3) they teach different courses. The students, who participated in the study were either in their first, second, or third year of studies.

2.2 Educational Context

Problem-based learning. The research was carried out at an institution of post-secondary education that organizes its curriculum according to principles of problem-based learning. Each class has about 20-25 students, guided by a tutor. Students work collaboratively in teams of 4-5, with learning centred on problems relevant to their domain of study. They work each day on one problem. The problem is initially discussed in the morning; followed by ample self-study. At the end of the day, information gathered is shared and elaborated upon. All students enrolled at the institution attend a different module everyday during a five-day work week, and take no more than five modules per semester. Each semester is 16 weeks (Alwis, 2007).

Assessment in the curriculum. The daily assessment approach consists of four, independent elements: (1) a self-assessment activity, (2) a peer assessment activity, (3) a reflection journal, and (4) a judgment by the tutor on how well students have performed during the day. Students are also tested on their acquisition of knowledge.

The self-assessment activity requires students to respond to 8 items inquiring about the quality of students' performance within their team, such as expressing ideas openly, pointing out agreements or disagreements of ideas, asking questions when in doubt, valuing the contributions of their peers, and so forth. The peer assessment activity involves students responding to 4 items inquiring about the cooperativeness and quality of contributions of peers within the team. Students are asked to respond to these items on a Likert five-point scale ranging from "strongly agree", "disagree", "neutral" and "agree" to "strongly agree." (See Appendix A).

The self- and peer assessment activities are formative as they train students to make judgments about aspects of their own learning and that of their peers' and to further improve on. They are also summative as students' self- and peer ratings are criteria for daily grading. All the assessment tools are conducted online and serve both formative and summative functions.

The reflection journal consists of a short essay created by the student, that is 'personal' and records his or her reflections of his or her learning process in respond to a journal question given by the tutor, which generally asked students to be reflective about their learning and development. Examples include "What are some of my learning strengths? What are some of my learning weaknesses? How well did I contribute to the teamwork?" Tutors also asked students to be reflective about subject-matter knowledge, and examples of tutor-asked journal questions include "What have I learnt today? How can I apply what I had learnt?" In general, the didactic purpose of the reflection journal is to encourage and record reflection in learning.

The tutor judgment consists of tutors' observations of students' learning (such as self-directedness, level of participation and quality contributions during the day, teamwork and communication skills) and tutor feedback to students. It serves both formative and summative purposes. Tutor judgment is formative as tutors will feedback on their observations of students' performance during the day and summative because it is a criterion for grading.

2.3 Procedure

A total of ten focus-group sessions were conducted. The focus group conducted at any one time consisted of either tutors or students, and involved no more than three participants. The participants were asked about (1) what they see as the purpose or functions of the various assessment procedures, in particular the self- and peer assessment activities, reflection journal and tutor judgment, and (2) the actual use of the various daily assessments. A list of questions (with several sub-questions) was used to guide the interview (see Appendices B and C). Each session lasted between 30 to 40 minutes. The sessions were recorded using recording software installed on a laptop. The data collected were analyzed and transcribed in the manner described in a study by O'Connell and Dymont (2006). The recordings were listened several times with a view of performing content analyses on the data. The raw data collected from the ten focus groups were transcribed, following which trends and patterns in the participants' responses that reappeared within a single focus group or among various focus groups were classified together. All evaluative remarks with regards to the different assessment tools were tallied and tabulate. Participants' names were omitted to protect their anonymity.

3. RESULTS

Table 1 contains the purposes of about the different elements of daily assessment reported in the literature, and as obtained from content analyses of the outcomes of the focus-groups interviews with tutors and students. Table 2 contains a summary of the perceptions of tutors and students about the use of the various assessments, and their experiences with each.

Table 1. Summary of the purposes of the different daily assessment tools: literature versus findings from focus-groups with tutors and students

Tool	Purpose (according to the literature)	Purpose (according to tutors)	Purpose (according to students)
Self-assessment	To (1) equip students with the ability to accurately assess one's strengths and weaknesses, an ability that is critical to the enterprise of lifelong learning (Boud & Falchikov, 1989; Eva et al., 2004; Gordon, 1992); (2) equip them with higher order thinking skills important for professional functioning (Sluijsmans et al., 2001)	"Students look into a mirror and form a complete picture of how they had performed for the day. My perception is for students to evaluate their performance but they might not see it that way." (T6+1) "To train students to reflect and to engender learning behaviours in them." (T2) "For students to evaluate aspects of their day's performance" (T7)	"It's about reflecting on our learning..., for tutors to look at how we evaluate ourselves so that they can use it to grade us." (S1+2) "For me to show my level of involvement during the day." (S5+2) "To let tutors know what happened when they are not in class..., help them to grade." (S11+4)
Peer assessment	To train students to be (1) self-regulated learners by which they can monitor their work using internal and external feedback such as peers' contributions in collaborative groups as catalysts (Butler & Winne, 1995) To increase learners' (2) understanding in the cognitive and metacognitive domains and to develop social and interpersonal skills (Topping, 1998) To train students to (3) reflect on performance of their peers; (4) equip them with higher order thinking skills important for professional functioning (Papinczak et al., 2007)	"(...) feedback on team dynamics and teamwork." (T4+1) "Reflection on teamwork and performance." (T3+1) "For students to evaluate their peers' contributions to the group." (T1+1)	"To reflect on the performance of my team mates and to help them to improve on their weaknesses." (S2+1) "To evaluate my teammates' performance." (S5+4) "To feedback to my tutor if my team mates had been performing or slacking." (S4+1) "Teammates can evaluate others so that tutor roughly knows who did what, how well each teammate performs." (S9+2)

Tool	Purpose (according to the literature)	Purpose (according to tutors)	Purpose (according to students)
Reflection journal	To (1) encourage and (2) record reflection in learning (O'Connell & Dymont, 2006; Wong et al., 1995) To (3) enable exploration of connections between ideas encountered on the course and the writer's experience (Creme, 2005; Grant et al., 2006; Langer, 2002; Thorpe, 2004)To (4) stimulate thinking and assist students in developing writing skills (Kerka, 1996)	"Trigger for students to reflect on certain aspect of the process or concepts that otherwise would not have been covered during the day." (T3+2) "For students to reflect on the content, the learning process and teamwork." (T7+2) "Channel for the students to talk about what they had done and felt about the day that happened." (T4) "Encourage students to reflect on their learning." (T6+2)	"To reflect on the things that I have learnt throughout the lesson so that the tutor can understand what I've learnt, what I don't know, what I misunderstood..., it's a good way to communicate with the tutor." (S1+6) "To reflect on my learning..., I get to recall what I learn for the whole lesson." (S4+5) "Helps me to reflect and also to improve on my writing skills." (S13+4)
Tutor judgment	To highlight (1) students' strengths and weaknesses in learning, and (2) suggestions for improvements in learning and performance; (3) motivator of new learning (Falchikov, 2005)To provide for (4) the consolidation of cognitive change, the deepening of understanding and realignment of concepts within each individual student's conceptual framework (Taras, 2001, 2002)	"(...) based on my observations, I tell my students how to improve, what they did well..., give them some indications of what I observe in class ..., what aspects of their performance were being emphasized." (T1+2) Highlight gaps or areas of weakness especially with regard to the use of resources or the learning process (T3+2) "To highlight issues about students' performance." (T3+4) Justification on why a particular daily grade was awarded (T1+2)	"Tutors give advice on what and how to improve in terms of performance, participation in class, and learning. (S4+4) "(...) for interaction, improve student-tutor relationship..., so I know why I am getting low grades..., help me to improve." (S3+3) "To motivate me... to highlight what I did during the day (good and bad) and how I can improve next week."(S8+2) "To encourage me to work harder." (S10+2)

Note: 'T' denotes tutors (T1-T7) and 'S' denotes students (S1-S15). (Tx+y) means that the finding was reported by tutor x and y tutors shared similar views. Students' perceptions were also reported in this way.

Table 2. Issues with daily assessment tools: a summary of tutors' and students' perceptions

Tool	Issue	Tutors' perceptions	Students' perceptions
Self-assessment	Are students overwhelmed by the daily rigor of assessing themselves?	<ul style="list-style-type: none"> • Not used effectively to aid students in their learning; students go through the rigorous process of evaluating themselves day after day causing it to lose its task value. (T2+3) 	<ul style="list-style-type: none"> ▪ "I don't read the statements, I just rate. I give the same answers every week, everyday." (S5+2) ▪ "The self-assessment is something I have to do every day; ...It's a waste of time and it doesn't help in my learning..."(S9+4) ▪ "Some of my friends just can't be bothered, just don't do the self-assessment or just give the same ratings."(S3+4) ▪ "For me it's quite redundant... for the last 3 years I rate myself the same."(S12+9) ▪ "I just do it blindly. Usually there are trends in the ratings for my self-assessment; it's usually all "agree" or all "neutral". (S7+5)
	Is the self-assessment a learning aid?	<ul style="list-style-type: none"> ▪ "I don't consider it as a reliable measure of anything... it indicates to me nothing much, I mean at the best, maybe when students are having problems I can get something out of self-assessment...but on a 'normal' day it indicates to me almost nothing." (T3+2) ▪ "(...) I mean do students really look upon how they evaluate themselves this week and last week... do they look back and improve on areas that they marked themselves down the previous week and try to improve? I mean, I don't really observe any 	<ul style="list-style-type: none"> ▪ "Actually the self-assessment does not really help me in my learning."(S2+7) ▪ "It is like a trigger for me to take work harder... when I reflect on how I performed I feel guilty that I didn't do much work."(S7+1)

Tool	Issue	Tutors' perceptions	Students' perceptions
		differences... I can't discern how students use the self-assessment to bring about some changes in their performance." (T4+4)	
	Should the self-assessment be used as a summative assessment tool?	<ul style="list-style-type: none"> ▪ "(...) it is not an assessment tool... it does not help to evaluate the students, maybe it can be used just as an indicator."(T3+2) ▪ "I don't use the self-assessment in isolation but together with the peer evaluation when I grade, so that I can countercheck students' self assessments against those by their peers." (T6+5) 	<ul style="list-style-type: none"> ▪ "Because for the tutor, they are not with us most of the time, so they can use the self-assessment to justify their grades." (S1+4) ▪ "I mean, tutors look at the self-assessment, but how true is are they?" (S2+3) ▪ "Who would want to rate themselves negatively?" (S7+3)
Peer Assessment	Are students reporting the truth about their peers' performances?	<ul style="list-style-type: none"> ▪ "Students are aware that their performance is being commented by others, in fact it keeps them honest."(T3) ▪ "(...)sometimes I just have to speak to the students... because I can see some students who are not doing much for the day but they still get "Agree" and "strongly agree" peer ratings... students just misuse the peer assessment." (T7+2) ▪ "I think my students misuse it to some extent. Sometimes I notice that some students are not around but are marked well by their team mates... to students, they feel that they need to be protective over their friends, and to them rating their team mates well as demonstrating team spirit." (T5+2) ▪ "Provides a channel for students to feedback on 	<ul style="list-style-type: none"> ▪ "Actually, there is a secret agreement between students. We usually tell each other that if you grade me well, I will grade you well... but this is not always so but most of the time; but then he/she still has to do some work." (S1+2) ▪ "Usually I don't mark below "Neutral" for my peer assessment; I try to help my friends." (S2+6) ▪ "Not so strict with peer ratings... tend to be more lenient, but if team mates are slacking, then my ratings will not go above neutral." (S3+7) ▪ "I don't think my peers take the peer assessment seriously, they just give the same ratings every day. Actually I also rate

Tool	Issue	Tutors' perceptions	Students' perceptions
		<p>team mates. When I see someone getting "strongly disagree" for peer assessment, then I know that there are some problems with the team." (T6+2)</p>	<p>all "agree" in my peer assessments. "(S4+5)</p> <ul style="list-style-type: none"> ▪ "If I want to score an A I don't want to commend my friends... since it's so competitive here... I don't want it affect my grades." (S7+3)
	<p>Are students overwhelmed by the daily rigor of assessing their peers?</p>	<ul style="list-style-type: none"> ▪ "It has a certain amount of rigor as the self-assessment, but most of the students use it to point out "slackers" in their teams." (T7+3) 	<ul style="list-style-type: none"> ▪ "Every day I am doing the same thing..., getting sick of it. Even if the team mates not doing anything I also don't bother about it." (S11+8) ▪ "(...) most of my peers just do the peer assessment for the sake of doing... they give bad ratings if they don't like their team mates." (S4+3)
	<p>Should the peer assessment be used as a summative assessment tool?</p>	<ul style="list-style-type: none"> ▪ "Highly useful for me... highly useful when students are frank with their evaluations."(T3+1) ▪ "...tally the ratings with the observations of students' learning... cross-validate the ratings with my observations of how students performed." (T4+2) ▪ "(...) from the peer assessment at least we can see if somebody is not contributing... it is more of a negative thing than a positive thing... we cannot really tell if students have contributed, but we can tell if they never do anything." (T2) ▪ ". Actually we face the problems of students as "actors", they behave differently when the tutor is 	<ul style="list-style-type: none"> ▪ "(...) it's just to tell my tutors whether my team mates do work or not." (S8+4) ▪ "We don't take it seriously, so the best is don't look at the peer assessment when tutors grade." (S12+5)

Tool	Issue	Tutors' perceptions	Students' perceptions
		in class and pretend to contribute and limited contact time with students... useful to help me to grade students." (T7+2)	
Reflection journal	Do journal questions which asked students to be reflective about what they had learnt (i.e. content) interfere with the didactic purpose of the reflection journal to encourage reflection on the learning process?	<ul style="list-style-type: none"> ▪ "I will ask students how they approach the problem and how they work among their teams, sometimes I will ask them to look at what they had done for the day ...have them looked into their problem-solving skills and strategies." (T5+2) ▪ "Issue is that some students assume that tutors are using the reflection journal to understand how much they learnt through the day..., I see some other tutors have graded students based on how much content they write in their reflection journals and they get good marks... this is bad, reflection journal should be used to encourage reflection on the process." (T3+2) ▪ "Even when I asked my students to reflect on their learning process, they still write a summary of what they had learnt for the day in addition to their reflection journal response to my question." (T7+2) 	<ul style="list-style-type: none"> ▪ "(...) sometimes the reflection journal question is related to today's problem then that will help us better understand what we learnt..., we don't get to tell our tutors what we understand so through the RJ they can know if we understand what we learnt." (S7+3) ▪ "The value of doing the reflection journal is dependent on the reflection journal question. I think I prefer content-based question, because answering it is like a revision of what I learnt for the day." (S11+4)
	Can reflection journal be used as a communication channel?	<ul style="list-style-type: none"> ▪ "The reflection journal is like a communication channel which students use to highlight problems on learning and teamwork." (T4+3) 	<ul style="list-style-type: none"> ▪ "It is something for me to communicate with the tutor, because they will ask me to reflect on my team mates, on my learning." (S5+4) ▪ The tutor is not in class all the time, so through the reflection journal, tutor can understand how much we learn and how

Tool	Issue	Tutors' perceptions	Students' perceptions
			<p>we did in class." (S9+3)</p> <ul style="list-style-type: none"> ▪ "Sometimes when I am very unhappy with my teammates, I will write to my tutor to request to change team." (S9+4) ▪ "If I don't really understand about the today's problem, I will ask my tutor in my reflection journal ... it's better to write in the reflection journal because no one can see it." (S1+3)
	Is the reflection journal a learning aid or a study aid?	<ul style="list-style-type: none"> ▪ "Usually students will reflect on the content first, then on their learning." (T6+2) 	<ul style="list-style-type: none"> ▪ "To some extent, it helps me to learn. I have to go through the notes that I take down during the tutors' presentation when I do my reflection journal." (S2+6) ▪ "Sometimes I am not really aware of my weaknesses, but when I do the reflection journal I become more aware of what I did not do so well." (S4+4)
	Should the reflection journal be assessed?	<ul style="list-style-type: none"> ▪ "I am not too sure if it should be part of the daily grade... it's like a "meditation"... soul searching... students should be encouraged to reflect but there is no need to make it part of the grade." (T5+4) ▪ "I will grade students based on how much content they write in their reflection journal ..., this gives an indication of how well and how much students have reflected on what they learnt for the day" (T6+5) 	<ul style="list-style-type: none"> ▪ "(...) if I don't do my reflection journal I cannot get an A grade, even if I participate in class and I don't submit my reflection journal I cannot get an A grade." (S7+6) ▪ "It does help to improve my grade by doing a reasonably good reflection journal, especially when I don't participate that well during the day."(S4+6) ▪ "The reflection journal is so important, it's even more important than punctuality for classes."(S5)

Tool	Issue	Tutors' perceptions	Students' perceptions
Tutor judgment	Are casual observations in the classroom sufficient to arrive at informed judgment?	<ul style="list-style-type: none"> ▪ "...very challenging because students behave differently when the tutors are in class." (T3+2) ▪ "When I observe how my students learn, I might miss out on something since I am not in class all the time and I have to deal with 25 students. Therefore, I use the self-evaluation and peer-evaluation to corroborate against my observations and I also look at their reflection journals before giving a final daily grade." (T4+5) ▪ "Observations of student learning is the best judge of the process... daily grade can be given based on the observations and the rest of the evaluation tools can be used to fill the gaps in what I observe." (T7+5) ▪ "It is quite difficult to be absolutely sure of the students' true performance... quiet students may suffer in our problem-based learning environment because they fear expressing their ideas openly... we cannot just depend on one assessment tool to grade students." (T5+2) 	<ul style="list-style-type: none"> ▪ "Of course we have to behave in the way to please the tutors so... we don't have a choice... the ultimate goal of coming to school is to get an A." (S1+1) ▪ "It's not so much of pleasing the tutors, but it's more about meeting the expectations of the tutors."(S3+1) ▪ "(...) it's not fair for the tutors to judge us based on how we participate in class, we don't speak up don't mean we don't understand... the best is still to judge us based on our reflection journals." (S14+5)
	Does tutor feedback motivate and improve learning?	<ul style="list-style-type: none"> ▪ "I follow up on my comments given in previous weeks... tell students when they have improved... encouraged them to work harder when I never observe any improvements in their performance." (T2+4) 	<ul style="list-style-type: none"> ▪ "Because only the tutors can see clearly how I perform, so if tutors don't tell me, I don't know what I should improve on." (S2+2) ▪ "...their words really play a part... encourage us to work harder..." (S7+3)

Note: 'T' denotes tutors (T1-T7) and 'S' denotes students (S1-S15). (Tx+y) means that the finding was reported by tutor x and y tutors shared similar views. Students' perceptions were also reported in this way.

4. DISCUSSION

The purpose of this study was to collect tutors' and students' perceptions of a daily assessment approach at an institution of post-secondary education using problem-based learning. Assessment is conducted daily and consists of four, independent elements: (1) a self-assessment activity, (2) a peer assessment activity, (3) a reflection journal, and (4) a judgment by the tutor on based on students' classroom performance. Focus-group interviews with tutors ($n_1=7$) and students ($n_2 =15$) were conducted and the collected data were qualitatively analyzed. The results of the analysis demonstrate that tutors and students have different views about the purposes and functions of the daily assessment approach.

It is clear from the responses tallied in Tables 1 and 2 that tutors and students differ to some extent in their perceptions of the daily assessment. Several issues arise from the similarities and differences in the beliefs, feelings, and experiences of tutors and students and these issues will be further discussed here.

Overall, there was congruence between the purpose of the various assessment tools according to the literature, and the understanding of both tutors and students about their purposes. For instance, the didactic purposes of the reflection journal reported in the literature are: (1) to encourage and (2) record reflection in learning; to (3) enable exploration of connections between ideas encountered on the course and the writer's experience; and to (4) stimulate thinking and assist students in developing writing skills. Both tutors and students mentioned these purposes several times when asked what they thought the intentions of the reflection journal were. To add on, the didactic purposes of the self-assessment reported by both tutors and students were congruent with what were reported in the literature, which is to equip students with the ability to accurately assess their strengths and weaknesses; and with cognitive skills such as reflection, problem solving and self-directedness.

A second observation is the following: according to the literature, the main purpose of journal writing is to get students to reflect critically on their learning *process* (Grant et al., 2006; Morrison, 1996).

Despite of this, students displayed a strong preference for reflecting on the *content* of the subject matter learnt rather than on the process. For example, the majority of the students perceived reflecting on the content when doing the reflection journal as a value-added task; they receive the opportunity (1) to revise what they learnt for the day; (2) to recall what they learnt for the whole lesson; and (3) to better understand the content of the subject matter acquired. It is clear that students' preference for reflecting on the content (rather than the process) contrasts sharply with the primary intention of the reflection journal, which is to get them to reflect critically on their process of learning. This finding is contrary to what was reported by Morrison, (1996):

"It is clear that for many students the 'openness to self' that the journal has fostered has contributed to what some of them refer to as the opportunity to recharge themselves (p.326)."

The third issue arising from the data collected is that students seem to use their self assessments and reflection journals not in the first place as instruments for the improvement of learning, but rather as tools for *impression management*. From the interviews it is clear that students put in a conscious attempt to make themselves 'look good' in front of their tutors through assessing their own performance better than they actually performed in their self assessments, and writing reflection journals that were quantitatively (if not qualitatively) good. Some examples of the verbatim responses made by students which substantiated this argument included: "Even if I participate in class and I don't submit my reflection journal, I cannot get an A grade." (S7+6), and "It does help to improve my grade by doing a reasonably good reflection journal." (S4+6). Students also expressed that they do not like to rate themselves negatively in their self assessments, quoting "it's just human nature (S7+3)" as the main reason for doing so.

It is clear from the responses tallied that students try to make a good impression on their tutors instead of using the self-assessment and reflection journal to demonstrate what they had learnt. Why is this so? In the particular curriculum studied it seems that the self-assessment and reflection journal are indirectly used in a summative fashion; that is: the measures themselves have a formative purpose, but tutors seem to consult students' self and peer assessments, and

their journal responses, as a corroboration of their impressions of students' performance in class to arrive at the daily grade. Because of the process of having to deal with 25 students in any lesson, tutors have limited time to observe each and every student's learning. As such, they may rely on other available sources of information such as students' self and peer assessments, and their journal responses, to arrive at the daily grade. Students being aware of the manner they are graded, adapt their behaviours accordingly through over-rating their own performance in their self assessments, and putting in conscious attempts to write quantitatively good (if not qualitatively) reflection journals. Some examples of the verbatim responses made by tutors: "When I observe how my students learn, I might miss out on something since I am not in class all the time and I have to deal with 25 students. Therefore, I use the self-evaluation and peer-evaluation to corroborate against my observations and I also look at their reflection journals before giving a final daily grade." (T4+5), "I will grade students based on how much content they write in their reflection journals ..., this gives an indication of how well and how much students have reflected on what they learnt for the day." (T6+5) and "I don't use the self-evaluation in isolation but together with the peer evaluation when I grade, so that I can countercheck students' self ratings against their peer evaluations." (T6+5). It may not come as a surprise that, under these circumstances, students respond by adapting their approach to this new use of the particular assessment tools. In fact, the reflection journal was identified as being most significant for the daily grades, whereas the other elements of daily assessment were perceived as less influential on the daily grade. This finding substantiates what Paterson (1995) reports about the problem of marking something that was expressly written for what the teacher wants to see, because "The student is generally preoccupied with what it is that the teacher wants in the response (p.219)."

A fourth issue raised by both students and staff is that students are assessed daily based on: (1) their self assessments, (2) their peer assessments, (3) their journal responses, and (4) judgment by a tutor on how well they have performed during the day. Students indicate to be overwhelmed by the daily rigor of having to evaluate

themselves and their peer's performance. They reported: assessing themselves "in the same way every day (S5+2)," and did their self-evaluation "blindly." They also remarked that the rigor of having to assess their own performance everyday was "too much" and "redundant (S7+5)." One student (S12) blatantly described "for the last 3 years I rate myself in the same way." This finding is consistent with what Eva et al. (2004) reports about students' ability to evaluate their own deficiencies did not improve with time in the program. Students also mentioned that they were "getting so sick of" having to evaluate their peers' performance everyday and mentioned that they were doing their peer assessments "only for the sake of doing it." Despite their peers contributing minimally towards teamwork, students reported that they could not care less to assess their peers accordingly.

The peer assessment is intended to measure students' collaborative skills, and to train students to (1) reflect on and (2) evaluate the performance of their peers. Although students see themselves as informers to feedback to tutors, peers who did not perform well during the day, they were generally lenient when evaluating their peers, and quoted "trying to help my friends" through "rating "agree" or "strongly agree" to the statements of the peer assessment", as the main reason for being less strict with peer ratings. Another issue which arises from both the rating errors associated with peer assessment as reported by Pond et al. (1995), and the daily rigor of the self- and peer assessment activities was the trends in students' self- and peer ratings. This is tricky because tutors become sceptical about using students' self and peer assessments to corroborate against their judgment of students' performance when grading.

In summary, both tutors and students were aware of the didactic purposes and actual functions of the different elements of daily assessment. Tutors use their judgments, students' self and peer assessments, and their journal responses in part when deriving the daily grade. In contrast, students' perceptions of daily assessment is less optimistic: (1) they prefer to reflect on the content rather than the process, (2) they used the self-assessment and reflection journal mainly as tools for impression management, viz., to make themselves

look good in front of their tutors; (3) they found the daily rigor of having to assess their own performance and that of their peers' every day overwhelming; and (4) they were less strict with peer assessments because they want to help their friends. As the study revealed, the value of the reflection journal lies in its perceived potential to reflect on personal learning achievements and to account for individual effort in a collaborative learning environment like problem-based learning. The other elements of daily assessment (i.e., self- and peer assessment, and tutor judgment) were perceived as less influential on the daily grade.

4.1 Implications

What are the implications of the present findings for new assessment practices in higher education? First, they suggest that continuous assessment if conducted too often may have a negative impact on student learning. As the study reveals, getting students to evaluate themselves and their peers' performance daily is akin to a "double-edged sword," which simultaneously helps students' improves on their learning and performance and that of their peers. Nonetheless, students are overwhelmed by the daily rigor of having to assess themselves and their peers.

Chapter 3- Measuring students' beliefs about self-assessment

ABSTRACT

The absence in the literature of a validated instrument to measure students' beliefs about the utilities of self-assessment practices was the motivation behind the present studies. To that end, a questionnaire was developed containing statements about the value of assessment procedures such as self-assessment tools and reflection journals. Students were able to identify the seven latent constructs underlying the questionnaire, as indicated by the fit of the hypothesized model. The test for measurement invariance showed that factor loadings were equivalent across different student groups and the questionnaire's underlying structure gave evidence of cross-validation. Evidence for sufficient test-retest reliability was also found suggesting stability of beliefs over time. These findings taken together demonstrate that the questionnaire developed appears to be an adequate instrument for measuring students' beliefs about the utilities of self-assessment. Factor correlations demonstrate that students believe that self-assessment can have multiple purposes, including self-improvement and impression management of teachers that are not necessarily in accordance with each other.

Keywords: *Self-regulated learning; students' beliefs; self-assessment; reflection journal; factor analysis*

1. INTRODUCTION

Research during the last 30 years on learning and achievement has progressively emphasized on cognitive strategies, metacognition, and authentic assessment practices used in the classroom. These various aspects of academic learning promote students' self-regulated learning. This form of learning emphasizes autonomy and control by the learner who monitors, directs, and regulates action towards goals of information acquisition, expanding expertise, and self-improvement (Paris & Paris, 2001). An example in which self-regulated theory has a direct application in the classroom is self-assessment. It is assumed that, through assessing their own performance, students engage in metacognition which enables them to reflect on their own accomplishments, to monitor their progress while learning, and to internalize standards of performance so that they can regulate their learning more effectively (Dochy, Segers, & Sluijsmans, 1999; Segers & Dochy, 2001).

Self-assessment is not only expected to encourage reflection or the self-appraisal of one's abilities, it is also supposed to actively engage students in their processes of learning. These features of students' learning are crucial in assisting them to become self-regulated, life-long learners who develop control over their own learning (Mok et. al, 2006; Paris & Paris, 2001). In addition, periodic self-assessment of learning processes and outcomes promotes monitoring of learning progress, and stimulates repair strategies which enable the learner to take further steps to improve on his learning deficiencies (Falchikov, 2005; Mok et al., 2006).

Many types of self-assessments are possible in the classroom. Students can reflect on their work and learn to assess their level of understanding, effort and strategies used on a task. They can also assess the improvements made in their learning over time (Eva et al., 2004). Journal writing also provides many opportunities for students to engage in self-assessment. Students reflect on their learning process and achievements and document these reflections, understand their learning progress through journal keeping, and share their personal responses to collaborative learning and school

work with their teachers (Chirema, 2007; Lew & Schmidt, 2006; O'Connell & Dymont, 2006).

In higher education in particular, there is a growing interest in alternative assessment practices which encourage students to become involved more actively in monitoring and reviewing their own performance. Examples of such practices are self-assessment tools and reflection journals, which also include techniques for peer assessment and portfolio assessment (Dochy et al., 1999; Falchikov, 2005; Paris & Paris, 2001). In this paper, we will focus on discussing relevant literature on self-assessment tools and reflection journals.

In the study by Dochy et al. (1999), they reviewed the literature on 63 studies from 1987-1998 on the use of self-, peer- and co-assessment in higher education. They indicated that students who engaged in self-assessment were more reflective and were better at problem solving. Similarly, Segers and Dochy (2001) explored undergraduate students' beliefs about self-assessment in a problem-based learning environment. Their findings revealed that students generally found the process of assessing themselves as stimulating deep-level learning and critical thinking. Journal writing has also been found to be of positive value to students. These benefits include the usefulness of the journal as a learning tool to encourage and record reflection in learning and to improve on writing skills (Kerka, 1996; O'Connell & Dymont, 2006). Through journal writing, students are suggested to become better aware of their learning accomplishments (Chirema, 2007).

By contrast, some other researchers are less optimistic about the use of self-assessment tools and reflection journals for improving students' learning. For instance, Maclellan (2001) conducted a qualitative exploration of teachers' and students' perceptions of assessment. Her study showed that although teachers maintained that they understood what the purpose of the various assessments was, self-assessment was infrequently used and, if so, exclusively at the end of a module. Teachers focused more on grading and ranking students instead of providing feedback on their learning. Students, on the other hand, reported that they did not exploit self-assessment to improve their learning and furthermore, appeared to have underdeveloped conceptions of what assessment was.

In a focus-group study, Lew and Schmidt (2006) compared teachers' and students' perceptions of assessing students' ability to reflect upon their learning. Their findings revealed that both teachers and students understood the purposes of the various assessment tools, but perceptions of their actual use differed. Teachers generally believed that self-reflection helped students to become better learners. On the contrary, students could not see the reflection journal and the self-assessment activity as valuable in their own right (as was the purpose). Students reported using the reflection journal as a study aid to summarize the contents of what they had learnt, rather than using it to reflect on their learning process. They also wrote about their peers' contributions as part of their journal responses. Furthermore, students believed that their self assessments and reflection journals could be used to influence the teachers' impressions of their performance, and were used by their teachers (to some extent) to arrive at the grades. Students also reported feeling overwhelmed by the daily rigor of journal writing and having to evaluate their own performance (in the particular curriculum students were required to reflect on a daily basis).

Langer (2002) investigated how adult learners used learning journals in a computer technology class. His study showed that students used the learning journal as a study aid to summarize the content of what they had learnt and in preparing for the examinations, rather than as a learning tool to reflect on their learning process. According to Langer, this interpretation by the students of what the learning journal was for negatively influenced the objective of developing critical thinking. In another study, Kerka (1996) also reported that students used the reflection journal as an instrument of attacking fellow students. She also highlighted the problem of marking journals expressly written for what the teachers wished to see, since the students were predominantly concerned with what their teachers wished to find in their journal responses.

In summary, there seem to be at least three problems associated with the use of reflection journals in higher education. First, students sometimes seem to use the reflection journal as a study aid to summarize the course materials, rather than using it as a tool to reflect on their own learning process. Second, students occasionally

seem to use the journal as an instrument to manage their teachers' impressions of them. Third, students sometimes seem to use the journal as a tool to criticize their fellow students.

Given the impact of any assessment on student learning, several researchers (e.g. Langer, 2002; Maclellan, 2001; Segers & Dochy, 2001) have emphasized the importance of striving to obtain a match between the didactic functions of the different self-assessment measures and how students make use of them in reality. The studies summarized here suggest that to fully exploit the benefits of assessment on students' learning, one must first understand how students *perceive* these measures. The way in which a student perceives assessment will determine the way he responds to it. To add on, the learner's experience of assessment determines the way in which the student tackles his learning. The present studies were conducted to investigate to what extent students believe that self-assessment contributes to improving their learning. Furthermore, the findings reviewed above were largely based on *ad hoc* questioning of students and staff about the utility of self-assessment. A calibrated instrument to measure students' beliefs about the utilities of such assessment tools is lacking.

In response to this omission, a 31-item questionnaire was developed containing statements about the utilities of assessment procedures such as self-assessment tools and reflection journals. In Study 1, a hypothesized model containing seven factors based on belief categories derived in the literature about student-reported views about the utilities of self-assessment and journal writing was developed. In Study 2, the resulting model of self-reflection beliefs was tested in a second, independent sample from the same student population to cross-validate the proposed model. In a third study, test-retest data were collected to study the stability or instability of the beliefs over time.

2. STUDY 1

Study 1 was conducted to validate a hypothesized questionnaire model containing seven factors based on belief categories derived

from the literature about the utilities of journal writing and self-assessment on students' learning.

2.1 Method

2.1.1 Subjects

Participants were 327 students in their second year of studies in the academic year 2007-2008. Second-year students were selected for the study because they were familiar with the daily assessment system having gone through 2 semesters (32 weeks) of studies. Of these 327 students, 171 (52%) were females and 156 (48%) were males. Their mean age was 18.80 years ($SD = 1.35$). The mean Grade Point Average (GPA) value of the participants at the end of the first semester of the academic year 2007-2008 was 2.84 ($SD = 0.53$). The GPA is calculated based students' classroom performance grades as awarded by their tutors, and their knowledge acquisition test grades. The participants were representative of the entire cohort of 2,573 second-year students (53% females, 47% males), with mean age of 18.83 years ($SD = 1.51$) and a mean GPA value of 2.80 ($SD = 0.53$).

2.1.2 Educational Context

Problem-based learning. The research was carried out at an institution of post-secondary education that organizes its curriculum according to principles of problem-based learning. Students work collaboratively in teams of 4-5, with learning centred on problems relevant to their domain of study. They work each day on one problem. The problem is initially discussed in the morning; followed by ample self-study. At the end of the day, information gathered is shared and elaborated upon. No didactic teaching takes place nor is there any other form of direct instruction. One tutor supervises five of the student teams in a larger classroom. His or her role is to facilitate students' learning (Alwis, 2007). All students enrolled attend a different module every day during a five-day work week, and take five modules per semester. There are two semesters in an academic

year (each semester is 16 weeks). All the curricula offered are three-year curricula.

Assessment in the curriculum. The daily assessment approach consists of four, independent elements: (1) a self-assessment activity, (2) a peer assessment activity, (3) a reflection journal, and (4) a judgment by the tutor on how well students have performed during the day. Students are also tested on their acquisition of knowledge.

The self-assessment activity requires students to respond to 8 items inquiring about the quality of students' performance within their team, such as expressing ideas openly, pointing out agreements or disagreements of ideas, asking questions when in doubt, valuing the contributions of their peers, and so forth. The peer assessment activity involves students responding to 4 items inquiring about the cooperativeness and quality of contributions of peers within the team. Students are asked to respond to these items on a Likert five-point scale ranging from "strongly agree", "disagree", "neutral" and "agree" to "strongly agree."

The reflection journal consists of a short essay created by the student, that is 'personal' and records his or her reflections of his or her learning process in respond to a journal question given by the tutor, which generally asked students to be reflective about their learning and development. Examples include "What are some of my learning strengths? What are some of my learning weaknesses? How well did I contribute to the teamwork?" Tutors also asked students to be reflective about subject-matter knowledge, and examples of tutor-asked journal questions include "What have I learnt today? How can I apply what I had learnt?" In general, the didactic purpose of the reflection journal is to encourage and record reflection in learning.

The tutor judgment consists of tutors' observations of students' learning (such as self-directedness, level of participation and quality contributions during the day, teamwork and communication skills) and tutor feedback to students. It serves both formative and summative purposes. Tutor judgment is formative as tutors will feedback on their observations of students' performance during the day and summative because it is a criterion for grading.

Students also need to take four knowledge acquisition tests per module, which are taken at different points during the semester. The duration of each test is 30 minutes and it consists of answering at least 3 structured questions. The tests are conducted in a supervised environment, similar to that of an end-of-course examination. Students are tested on their ability to understand and apply what they had learnt.

2.1.3 Instrument

The absence in the literature of a validated instrument to measure students' beliefs about the utilities of self-assessment measures was the motivation behind the development of the questionnaire. A 31-item questionnaire which contains statements inquiring about students' beliefs about the self-assessment measures such as reflection journals and self-assessment was developed. The items were rewritten and refined several times before they were pilot tested on 327 second- year students. Students were asked to respond to these items on a Likert five-point scale ranging from "strongly disagree," via "disagree," "neutral" and "agree" to "strongly agree."

The questionnaire was designed to measure seven belief categories derived from past research studies on the use of the reflection journals and self-assessment tools in higher education. The latent constructs of the questionnaire model were described by means of statements instead of phrases. It was more informative to use statements, as phrases circumscribed the description of the belief category from which the construct was derived.

The self-assessment section consisted of 13 items and three underlying constructs. A study by Mok et al. (2006) revealed that students became better aware of their learning and thinking process through engaging in self-assessment, thereby enabling them to take steps to improve on their learning deficiencies. Items such as "The self-assessment helps me to assess my strengths and weaknesses accurately" served to measure the construct on the usefulness of self-assessment in appraising students' learning. In another study, Lew and Schmidt (2006) reported that students also seem to use the self-

assessment as an impression management tool. The self-assessment impression management construct was measured by items such as "The self-assessment is mainly useful in managing the tutor's impression of my performance." They also highlighted the issue that students regarded assessing themselves more as a habitual action than as a means of improving their learning. This belief category was measured by items such as "I do the self-assessment without thinking how the statements are related to my performance during the day."

The reflection journal section consisted of 18 items and four underlying constructs. O'Connell and Dymont (2006) suggested the usefulness of the journal as a learning tool to encourage and record reflection in learning. An example of an item reflecting the usefulness of journal writing which helped students to think and write reflectively was "Writing the reflection journal enables me to explore what I have learnt in my modules and my own ideas about these subjects." In her work, Kerka (1996) proposed that the engaging in journal writing helped students to improve on their writing skills, and students wrote about their peers' contributions towards team work in their journal responses. The use of the reflection journal by students as a feedback channel to their teachers about teamwork was represented by items such as "I write about the contributions of my team mates in my reflection journal."

In their study, Lew and Schmidt (2006) also reported that students used their journals as impression management tools (to manage their teachers' impressions of them). Items reflecting the usefulness of the reflection journal in managing teachers' impressions of students' performance were for example "I can make myself look good in front of my tutor through writing a qualitatively good reflection journal." They also suggested that frequent journal writing improved students' learning. The construct about frequent journal writing improved students' learning was represented by items such as "Writing the reflection journal changes the way I learn."

2.1.4 Procedure

The questionnaire was administered online to the participants of Study 1 at the start of the 2007-2008 academic year. The questionnaire's instruction stated that there were no right or wrong answers to the items, all answers were correct so long as they reflected students' opinions. No information was given regarding the constructs underlying the questionnaire. Filling in the questionnaire took approximately three to five minutes.

2.1.5 Analysis

Data from Study 1 were analyzed using a structural equation modelling approach to test whether the underlying structure fitted the belief categories derived from the literature. The results showed a fairly poor fit between the seven constructs and their items: $\chi^2(436, n_1 = 327) = 1247, p < .00$, Bentler's comparative fit index (CFI) = .82, the root mean square error of approximation (RMSEA) = .08 (explanation of these indices will follow later in this section). To maximize the fit between the items and their underlying constructs, the model was modified by examination of the item's modification indices (Byrne, 2001). This exploratory analysis resulted in the elimination of 6 items with high modification index values. The deleted items often let room for multiple interpretations, or were mere replications of other items. The revised questionnaire containing 25 items was called the Self-Assessment Questionnaire (SAQ). The resulting questionnaire model with seven factors was tested with confirmatory factor analysis. Parameters for the model specified were generated using maximum likelihood. For the evaluation of the proposed model presented in this study, two groups of fit indices, absolute and incremental, were selected.

In this study, χ^2 , accompanied by degrees of freedom, sample size and p -value, as well as the root mean square error of approximation (RMSEA) were used as absolute fit indices. The χ^2 is used to test the goodness-of-fit between an observed and predicted covariance matrix. A small χ^2 value, relative to the degrees of freedom (i.e. ≤ 3) is an indication of good fit and vice versa (Byrne, 2001). RMSEA appears

to be sensitive to model specification, minimally influenced by sample size, and not overly influenced by estimation method and was therefore included. The lower the value of RMSEA, the better the fit, with a cut-off close to .06 (Hu & Bentler, 1999). The incremental fit index included in our study was the comparative fit index (CFI). It ranges from 0 to 1, with higher values indicating a better fit. Values greater than .90 are traditionally associated with well-fitting models, although more recently, values close to .95 are suggested (Byrne, 2001; Hu & Bentler, 1999).

Coefficient H values for each construct were also computed to determine construct reliability. Unlike other measures of construct reliability (such as Cronbach's alpha), coefficient H is unaffected by the sign of standardized factor loadings, and a value of at least .80 is considered reliable for a construct. This is because the higher the values of coefficient H would mean stronger and more stable factor loadings which tend to fluctuate less from sample to sample (Hancock & Mueller, 2001). In addition, unlike Cronbach's alpha, coefficient H is not sensitive to number of items included in a scale.

2.2 Results

Table 1 contains the regression weights and reliability coefficients H of the 25 items of the SAQ for Studies 1 and 2.

Table 1. Standardized regression weights (SRWs) and reliability coefficients H for the 25 individual items of the self-assessment questionnaire (SAQ) in studies 1 and 2

Tool	Construct	Item	Study 1		Study 2		
			SRWs	H	SRWs	H	
Self-assessment	The self-assessment enables me to make an appraisal of my learning.	1	Doing the self-assessment enables me to judge my performance better.	.85		.75	
		2	The self-assessment enables me to improve on my learning in areas that I'm not so good at.	.84		.78	
		3	I become better aware about my learning through doing the self-assessment.	.80	.95	.85	.94
		4	The self-assessment helps me to assess my strengths and weaknesses accurately.	.71		.73	
	Assessing my own performance is more of a habitual action than to improve on my learning.	5	Doing the self-assessment everyday is too frequent.	.74		.74	
		6	Doing the self-assessment is a waste of time.	.89	.93	.84	.89
		7	I do the self-assessment without thinking how the statements are related to my performance during the day.	.61		.65	
	The self-assessment enables me to manage my tutor's impressions of how I performed.	8	The self-assessment is mainly useful in managing the tutor's impression of my performance.	.80		.66	
		9	My tutor looks at my self-assessment when he/she grades.	.56	.86	.61	.84
		10	I assess myself in order for the tutor to grade me.	.19		.40	

Tool	Construct	Item	Study 1		Study 2		
			SRWs	<i>H</i>	SRWs	<i>H</i>	
Reflection journal	The reflection journal helps me learn to think and write reflectively.	11	Writing the reflection journal helps me to think critically about my learning.	.80		.78	.83
		12	Doing the reflection journal improves my writing skills.	.73		.67	
		13	Writing the reflection journal enables me to explore what I have learnt in my modules and my own ideas about these subjects.	.78	.94	.78	
		14	Because I have to write the reflection journal every day, I am better at expressing myself.	.74		.68	
	Frequent journal writing improves my learning.	15	The reflection journal is mainly useful for summarizing what I have learnt during the day.	.24		.61	
		16	Writing the reflection journal everyday is too frequent.	.54	.89	.67	.84
		17	Writing the reflection journal don't really change the way I learn.	.77		.42	
		18	If it was up to me, I prefer to write a reflection journal only once every week.	.78		.72	
	I can look good in front of my tutor when I write a qualitatively good reflection journal.	19	When I don't perform so well during the day, my reflection journal can help to improve my daily grade.	.72		.79	
		20	Writing a qualitatively good reflection journal improves my daily grade.	.61		.75	
		21	I can make myself look good in front of my tutor through writing a qualitatively good reflection journal.	.55	.89	.62	.92
		22	The reflection journal is mainly useful in that it helps the tutor to grade us.	.67		.69	
		23	When I write my reflection journal, I do my best to impress the tutor.	.47		.37	

Tool	Construct	Item	Study 1		Study 2		
			SRWs	<i>H</i>	SRWs	<i>H</i>	
Reflection journal	The reflection journal enables me to feedback to my tutor about my peer's performance.	24	When my team mates don't contribute, I feedback to my tutor in my reflection journal.	.61		.62	
		25	I write about the contributions of my team mates in my reflection journal.		.79		.82
			.69		.76		

Table 2 shows the correlations among the seven constructs of the SAQ for Study 1.

Table 2. Correlations among the seven constructs of the SAQ in Study 1

	1	2	3	4	5	6
1. The self-assessment enables me to make an appraisal of my learning.	---					
2. Assessing my own performance is more of a habitual action than to improve on my learning.	.56**	---				
3. The self-assessment enables me to manage my tutor's impressions of how I performed.	.59**	.75**	---			
4. The reflection journal helps me learn to think and write reflectively.	.57**	.33**	.42**	---		
5. Frequent journal writing improves my learning.	.35**	.27**	.32**	.72**	---	
6. I can look good in front of my tutors when I write a qualitatively good reflection journal.	.32**	.04	.26**	.47**	.25**	---
7. The reflection journal enables me to feedback to my tutor about my peers' performance.	.34**	.07	.27**	.35**	.21**	.43**

Note: ** Correlation is significant at the .01 level

2.3 Discussion

Analysis of the questionnaire model resulted in a CFI of .93 and RMSEA of .05. These values suggest a fairly good model fit (Byrne, 2001; Hu & Bentler, 1999). Results of the χ^2 analysis was $\chi^2(269, N = 327) = 480.22, p < .00$. The calculated constructs reliability values (coefficient H values) of the model range from .79 to .95, which reflected good construct reliability (Hancock & Mueller, 2001).

Table 1 shows the regression weights of the seven constructs. All loadings are significant at the .01 level, indicating that these items contributed significantly to their respective construct.

Correlations among the seven constructs are displayed in Table 2. The results suggest that the level of student belief about the usefulness of the self-assessment in appraising learning is positively associated with their beliefs that frequent self-assessment does not contribute to improving learning, and the use of the self-assessment as a tool to manage their tutors' impressions of them. With regards to the reflection journal, students who believed that the reflection journal helps them learn to think and write reflectively, also believed that their learning improves with frequent journal writing, the use of the reflection journal as an impression management tool, and as a means to feedback on teamwork.

In summary, a hypothesized model consisting of 25 items about students' beliefs about the utilities of self-assessment was tested. The results revealed a reasonable fit of the model with the data. This finding implies that students are able to distinguish between the different functions that these two assessment procedures are supposed to have and the actual role they may play, as reported in the literature.

3. STUDY 2

In seeking evidence for multigroup invariance, Study 2 was conducted to investigate if the hypothesized SAQ model replicates across independent samples of the same population. This addresses the issue of cross-validation (Byrne, 2001).

3.1 Method

3.1.1 Participants

Participants were 273 students in their second year of studies in the academic year 2007-2008. Of these 273 students, 143 (52%) were females and 130 (48%) were males. Their mean age was 19.03 years ($SD = 1.71$). The mean GPA value of the participants at the end of the

first semester of the academic year 2007-2008 was 2.83 ($SD = 0.51$). The participants were representative of the entire cohort of second-year students.

3.1.2 Educational context

Study 2 was conducted in the same institution as Study 1.

3.1.3 Instrument

The questionnaire derived from the Study 1 was used. Twenty-five statements had to be rated on a Likert five-point scale.

3.1.4 Procedure

The same instructions were given as in Study 1.

3.1.5 Analysis

The same analyses as in the Study 1 were applied to the data of this study. The χ^2 statistics CFI and RMSEA were used as fit indices.

3.2 Results

Table 3 shows the correlations among the seven constructs of the SAQ for Study 2.

Table 3. Correlations among the seven constructs of the SAQ in Study 2

	1	2	3	4	5	6
1. The self-assessment enables me to make an appraisal of my learning.	---					
2. Assessing my own performance is more of a habitual action than to improve on my learning.	.67**	---				
3. The self-assessment enables me to manage my tutor's impressions of how I performed.	.69**	.84**	---			
4. The reflection journal helps me learn to think and write reflectively.	.45**	.41**	.40**	---		
5. Frequent journal writing improves my learning.	.25**	.12*	.18**	.54**	---	
6. I can look good in front of my tutors when I write a qualitatively good reflection journal.	.33**	.17**	.29**	.44**	.24**	---
7. The reflection journal enables me to feedback to my tutor about my peers' performance.	.39**	.20**	.28**	.46**	.64**	.40**

Note: * Correlation is significant at the .005 level

** Correlation is significant at the .01 level

Table 4 gives a summary of the goodness-of-fit statistics for tests of invariance.

Table 4. Summary of goodness-of-fit statistics for tests of invariance

Model Description	χ^2	df	$\Delta\chi^2$	Δ df	Statistical Significance
Hypothesized model	1100.80	542	-	-	-
Model with factor loading constrained as equal	1118.30	557	16.50	15	NS

Note: χ^2 = Chi-square, $\Delta\chi^2$ = difference in chi-square values between the two models; Δ df = difference in degrees of freedom between models; NS = not significant at the .05 level.

3.3 Discussion

Similar to Study 1, maximum likelihood linear structural relations analyses were conducted. Results showed a CFI of .93 and a RMSEA of .05. Results of the χ^2 analysis was $\chi^2(268, N = 273) = 493.00, p < .00$. The calculated constructs reliability values (coefficient H values) of the model range from .81 to .94, which reflected good measurement reliability (Hancock & Mueller, 2001). The standardized regression weights for the individual items are contained in the second part of Table 1. All the loadings are significant at the .01 level, indicating that these items contributed significantly to their respective construct.

Table 3 shows the correlations among the seven constructs for Study 2. As with what was suggested by the findings from Study 1, the more students agreed that the self-assessment enables them to appraise their learning, the more they considered assessing themselves as a habitual action than to improve on their learning, the more they appeared to be using the self-assessment as a tool to manage their tutors' impressions of them, rather than use it to appraise their learning. Again, students who believed that their frequent journal writing enables them to think and write reflectively also regarded the reflection journal as a tool for impression management, and to feedback on teamwork.

In order to investigate if the questionnaire model presented in Figure 1 replicates across independent samples of the same population, test of measurement invariance was conducted across the two samples simultaneously. This gave rise to a CFI of .91 and a RMSEA of .04. These values are indicative that the SAQ model with seven constructs is fairly well fitting for the two student groups. The factor structure of SAQ repeats itself across the two independent samples, giving evidence of cross validation. The χ^2 value of 1100.80 ($N = 600$, i.e., the two student samples together) with 542 degrees of freedom provided the baseline value against which the model with equality constraints imposed was compared. This fixed factor model generated a CFI of .91 and a RMSEA of .04.

Table 4 contains the summary of the goodness-of-fit statistics between the hypothesized SAQ model and that of the fixed model.

The comparison of the χ^2 between the two models yields a difference value of 16.5, with 15 degrees of freedom, which is statistically insignificant at the .05 probability level. This result implies that the relations between the items and their latent constructs are equal across the different samples. Furthermore, the results suggest that the factor loadings were equivalent across the different samples and that the factor structure gave evidence of cross-validation.

4. STUDY 3

Study 3 was conducted to investigate if the validated SAQ from Studies 1 and 2 exhibits measurement stability across time. In doing so, the SAQ can be assessed for its test-retest reliability (Brace, Kemp, & Snelgar, 2006).

4.1 Method

4.1.1. Participants

Participants were 66 students in their second year of studies in the academic year 2007-2008. These students participated in the earlier studies. Of these 66 students, 35 (53%) were females and 31 (47%) were males. Their mean age was 18.92 years ($SD = 1.50$). The mean GPA value of the participants at the end of the first semester of the academic year 2007-2008 was 2.86 ($SD = 0.47$). The participants were representative of the entire cohort of second-year students.

4.1.2 Educational context

Study 3 was conducted in the same institution.

4.1.3 Instrument

The SAQ was used.

4.1.4 Procedure

The SAQ was administered to the participations on two occasions (once during Study 1 and the second time during Study 2) separated from one another by 10 weeks.

4.1.5 Analysis

Correlation coefficients between the seven constructs of the SAQ for Studies 1 and 2 are used as a quantitative measure of the test-retest reliability of the SAQ. Test-retest reliability involves testing the same participants with the same instrument on two separate occasions, and obtaining the correlation between the two sets of scores. Correlation values close to .70 are indicative of scale stability over time of a given instrument (Brace et al., 2006).

4.2 Results

Table 5 shows the correlation coefficients among the seven constructs of the SAQ for Studies 1 and 2.

Table 5. Test-retest reliability: correlation coefficients for given constructs between studies among the seven constructs of the SAQ in Study 3

Construct	Correlation coefficient
The self-assessment enables me to make an appraisal of my learning.	.74**
Assessing my own performance is more of a habitual action than to improve on my learning.	.69**
The self-assessment enables me to manage my tutor's impressions of how I performed.	.69**
The reflection journal enables me to feedback to my tutor about my peers' performance.	.73**
The reflection journal helps me learn to think and write reflectively.	.64**
Frequent journal writing improves my learning.	.63**
I can look good in front of my tutors when I write a qualitatively good reflection journal.	.66**

Note: ** Correlation is significant at the .01 level

4.3 Discussion

Table 5 provides the strength of relationship between the seven constructs of the SAQ for Studies 1 and 2. All the correlation coefficients, ranging from .63 to .74, are statistically significant at the .01 level. This indicates the test-retest reliability (measurement stability over time) of the SAQ (Brace et al., 2006). The results further suggest the stability over time of students' beliefs about the utilities of self-assessment and journal writing and on their learning. For instance, students who agreed that their self assessments enable them to make appraisals of their learning continued to think so despite a time lapse of 10 weeks. The SAQ scale on the usefulness of the self-assessment in appraising students' learning has a test-retest reliability value of .74.

5. GENERAL DISCUSSION

The present studies were conducted to investigate students' beliefs about the utilities of self-assessment. Seven belief categories derived from the literature about the value of assessment procedures such as reflection journals and self-assessment were included in a questionnaire. In the first study, the questionnaire was pilot tested on a group of students. The generalizability of the revised questionnaire from the first study was tested in a new, independent sample in the second study.

The results of these studies indicate that the questionnaire was able to measure students' beliefs about the value of assessment procedures such as reflection journals and self-assessment. Students were able to identify the different factors underlying the questionnaire distinctly, as indicated by the fit of the hypothesized model. Data from the two, independent student groups fitted this model fairly well. Results for the test of measurement invariance revealed that factor loadings were equivalent across the different samples, and that the questionnaire factor structure stood up to cross-validation. In addition, construct reliability values of the seven scales gave evidence of good reliability in terms of internal consistency. Furthermore, the test-retest reliability of the validated

instrument suggested its measurement stability over time. In summary, the questionnaire developed in appeared to be an adequate instrument to measure students' beliefs about self-assessment.

Factor correlations show that frequent journal writing is most positively associated with helping students to think and write reflectively. This indicates that the majority of the students believe that the more frequent they engage in journal writing, the better they are at critical thinking and reflective writing. To add on, students who believed that frequent journal writing helps them to learn to think and write reflectively also believed that the reflection journal can be used as a tool to manage their tutors' impressions, and to feedback about teamwork. With regards to the self-assessment, factor correlations indicate that students who believed that that it enables them to appraise their learning, also regarded assessing themselves more of as a habitual action than to improve on their learning, and think that the self-assessment can be used as a tool to manage their tutors' impressions of them. It is clear from these findings that students believe that self-assessment can serve multiple purposes, some of them not necessarily in accordance with each other.

5.1 Future Research

Two issues present themselves for further research based on the findings from these studies. First, the questionnaire model should be tested in independent student groups. This helps us to investigate if the items of the SAQ operate equivalently across different groups (e.g. age, gender, or experiences with journal writing and self-assessment). Second, further research should investigate the predictive validity of the questionnaire with respect to academic performance. If beliefs such as the ones measured with the present questionnaire about self-assessment contributing to learning, then their influence should be reflected in student achievement.

Chapter 4- Accuracy of students' self-assessment and their beliefs about its utility

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (in press) Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education*

ABSTRACT

The purpose of the two studies presented here was to evaluate the accuracy of students' self-assessment ability, to examine whether this ability improves over time, and to investigate whether self-assessment is more accurate if students believe that it contributes to improving learning. To that end, the accuracy of the self assessments of 3588 first-year students enrolled in a post-secondary institution was studied throughout a semester during which each student made approximately 80 self assessments about his or her own learning process. These self assessments were then compared with multiple judgments by peers and tutors. The overall correlations between the scores of self, peer and tutor assessments suggest weak to moderate accuracy of student self-assessment ability. The findings also reveal an ability effect; students judged as more academically competent were able to self-assess with higher accuracy than their less competent peers. Comparing the accuracy of student self-assessment averaged over four consecutive periods indicates that the accuracy does not improve over time. In a second study, a questionnaire aimed at eliciting student beliefs about the effects of self-assessment on their learning was administered to 936 first-year students. Based on their responses, sub-groups of students were identified: those who either believed in the usefulness of self-assessment or did not. Results suggest that there is no significant association between student beliefs about the utility of self-assessment and the accuracy of their self assessments.

Keywords: *self-assessment; self-assessment accuracy; peer assessment; tutor judgment; student beliefs*

1. INTRODUCTION

The upsurge of interest in student self-assessment among researchers and educators arises from the recognition of the possible positive role that self-assessment may play both in learning and in the development of professional competence (Boud, 1989; Sluijsmans, Moerkerke, & Dochy 1998). This interest must be seen in the light of the changing goals of higher education where the focus is no longer about just making students knowledgeable within their domains of study, but also to equip them with transferable skills for successful functioning in professional life (Dochy, Segers, & Sluijsmans 1999). The development of students' abilities to assess and evaluate their own work in ways applicable in their future profession is one such valuable skill (Stefani, 1994). It has been argued by some that self-assessment is a *sine qua non* for effective learning and is a critical tool for learning beyond university education (Black & William, 1998; Taras, 2001). Boud (1989) further argues that one of the defining characteristics of effective learners is that they have a realistic sense of their own strengths and weaknesses, and that they can use knowledge of their own achievements to direct their studying into productive directions. He goes on to emphasize that in the sphere of professional education, the need to monitor one's own performance is one of the defining characteristics of professional work.

Self-assessment has been associated with moves towards developing greater student autonomy and responsibility in learning, particularly self-regulated learning. Numerous authors (e.g., Paris & Cunningham, 1996; Paris & Paris, 2001) support the premise that processes of self-regulated learning enable the learner to monitor, direct and regulate his actions towards goals of information acquisition, expanding expertise and self-improvement. One of the critical self-regulatory skills that students need is the ability to self-assess. It is hoped that through self-assessment, students can internalize standards of professional expertise and reflect on their progress, enabling them to regulate their learning more effectively.

Self-assessment has also been linked to authentic assessment and the development of metacognitive skills. According to

Kraayenoord and Paris (1997), one of the main purposes of authentic assessment is to encourage students to become involved more actively in monitoring and reviewing their own performance. They also emphasize the central role of self-assessment in authentic assessment. Here, self-assessment is the key aspect of the evaluation of the products as well as the processes of daily learning so that students learn to reflect on their work and to evaluate their effort, feelings and accomplishments, not just their past grades. Kraayenoord and Paris also stress that as self-assessment includes both reflection and evaluation of one's work, it helps to develop responsible and autonomous learners who are capable of regulating their own learning. These features of students' learning are considered important in assisting them to become independent learners who are capable of controlling their performance and processes of learning.

A particular emphasis on metacognitive skills is evident in the definitions of and research on student self-assessment. Vockell (2004) describes metacognitive skills as the learners' automatic awareness of their own knowledge and their ability to understand, control and manipulate their own cognitive processes. In reviewing the literature in the past century on teaching and learning, the American Psychological Association (1997) highlighted metacognition as one of the more important factors in contributing towards effective learning. The review suggests that as students' metacognitive abilities develop, so do their abilities for self-reflection and self-regulation of learning, which in turn will lead to improvements in academic performance. This is illustrated in the work by Lopez and Kossack (2007), whose study explores the effects on student perceptions and academic performance when self-assessment was required several times throughout the course for some students but not for others. Their findings indicate that the course grades for students who used continuous self-assessment showed a more consistent increase across the course tests and were higher than the test scores of students who only engaged in pre- and post course self-assessment, or used no means of self-assessment at all. End-of-course correlations between students' self assessments and actual course grades were more significantly aligned for the

continuous self-assessment group, suggesting that students were more realistically aware of their abilities when they periodically evaluate their understanding of course knowledge. Furthermore, students remarked that they placed a higher emphasis on the nature of, and responsibility for, their own learning when self-assessment occurred throughout the semester. In another study, Mok and others (2006) also suggest that the use of a metacognitive approach for self-assessment of teacher education students enhances the learners' awareness of their thinking processes and learning. Subgroups of students were asked to represent their learning using concept maps, and were interviewed about their experiences with self-assessment. Analyses of the concept maps drawn by participants at the end of learning contained significantly more concepts and relationships than those drawn at the start of learning. Interviews with students revealed that they found the metacognitive approach supportive of their learning. Based on these findings, Mok et al. (2006) contended that such an approach for self-assessment might have led to changes in students' metacognition and processes of knowledge construction. However in this study, a control group was missing.

Positive findings with regard to the use of student self-assessment in classrooms have been reported in the literature. For instance, Dochy et al. (1999) analyzed 63 studies published between 1987-1998 which investigated the use of self-, peer and co-assessment in higher education. Their review suggested that the use of self-assessment in educational practice is useful for improving student learning. They reported that self-assessment, when used to promote learning of skills and abilities led to more responsible learners, demonstrating increased reflection of one's own work and had better problem-solving skills. Similarly, Segers and Dochy (2001) explored undergraduate students' beliefs about self-assessment in a problem-based learning environment. Their findings revealed that students generally found the process of assessing themselves as stimulating deep-level learning and critical thinking. Thompson, Pilgrim, and Oliver (2005) go on to show that self-assessment when used to encourage them to reflect more on what and how they learnt was to some extent, effective in developing critical thinking skills in students.

By contrast, other researchers are less optimistic about the effects of self-assessment on student learning. In a focus-group study, Lew and Schmidt (2006) compared teachers' and students' perceptions of the use of self-assessment. Their findings suggested that both teachers and students understood the purposes of self-assessment but perceptions of its actual use differed. Teachers generally believed that self-reflection helped students to become better learners. By contrast, students could not see the self-assessment as valuable in its own right (as was the purpose). Many of them believed that their self assessments could be used to influence the teachers' impressions of their performance, and were used by their teachers (to some extent) to arrive at their final grades. In another study, Maguire, Evans, and Dyas (2001) demonstrated how first-year university students when presented with self-assessment tasks, became 'strategic' in their approach to completing the tasks. In other words, the study by Maguire and his co-workers shows how students were able to spot the possibility of achieving good results with minimal work and took advantage of that. Students were sceptical about self-assessment and reflection, citing them as 'mechanical, meaningless tasks' which were of no benefit to improving their learning.

Besides the effects of self-assessment on student learning reviewed above, there is a body of literature reporting empirical studies that compare student-provided marks with those of teachers. In light of this type of self-assessment, research usually looks into the validity of the grades, by comparing the accuracy of the grade given by the learner with that given by the teacher or their peers (Boud & Falchikov, 1989; Falchikov & Boud, 1989). For instance, Cassidy (2007) asked 160 first year undergraduate students from the faculty of health sciences to provide marks for their own work, which were then compared with tutor marks. Teachers and students differed significantly in their judgment; a significantly positive but fairly low correlation of .25 existed between tutor and student-estimated marks. In addition, there was a higher tendency for students to underestimate (56%) than to overestimate (40%) their assignment marks. The investigator concluded that the majority of the students demonstrated a good level of self-assessment skill, with a quarter of the students failing to demonstrate such skills.

Some researchers have investigated the accuracy of student self assessments over time. For instance, Fitzgerald, White, and Gruppen (2003) conducted a longitudinal study which examined the ability of medical students to predict their examination performances accurately during their first three years in medical school. Correlations between students' estimated self assessments on knowledge examinations and their actual examination performance in their first two years of medical school were moderately high, ranging from .46 to .69. The correlations between students' estimated self assessments on clinical examinations and their actual examination performance in their third year were lower, ranging from .12 to .42. The results indicate that students' ability to self-assess their knowledge accurately was higher compared with their ability to self-assess their clinical skills. The findings also suggest that self-assessment might be influenced by task familiarity: when the task was one in which students had limited experience, self-assessment accuracy suffered, as did performance.

Besides self-assessment, peer assessment has also received much attention in higher education. Boud (1986) suggests that the contribution of other students can be a very useful input into the self-assessment process. Learners have an opportunity to observe their peers throughout the learning process and often have a more detailed knowledge of the work of others than do their teachers. Boud further contends that peer assessment should not be considered only as a grading procedure; it should be seen as part of the self-assessment process and which "serves to inform self-assessment" (p. 22). Many researchers (e.g., Nicol & Milligan, 2006; Segers & Dochy, 2001) have argued that through reflecting on and evaluating the performance of their peers, students can develop objectivity in relation to standards which can be then be transferred to their own work and real-world professional settings. Race (1998) goes on to say that as peer assessment requires students to use their knowledge and skills to review, clarify and correct others' work, this in turn would enhanced the metacognitive understanding about their own learning process (see also Ballantyne, Hughes, and Mylonas (2002)). Additional benefits of peer assessment suggested in the literature include improved understanding of subject matter, and

encouraged the development of skills of reflection and critical reasoning skills (McDowell, 1995; Searby & Ewers, 1997). Although strong support for peer assessment is evident in the literature, difficulties and limitations have repeatedly been reported. Students' perceptions that peer assessment can be unreliable and unfair have been raised in several studies in higher education (e.g., Dochy, Segers, & Sluijsmans, 1999; Lew & Schmidt, 2006; McDowell, 1995).

As with self-assessment, the existing literature on peer assessment has been dominated by empirical studies, which examined the accuracy of peer awarded marks as compared to student or tutor awarded marks. For instance, a qualitative review by Topping (1998) that focused primarily on the mechanisms and benefits of peer assessment located 25 studies, which compared teacher and peer marks, and 8 studies, which compared student and peer marks. The majority of the studies (18) demonstrate that peer assessments were of adequate reliability when compared to student and teacher marks in a variety of applications. However, Topping's study gives no indication of any discipline or subject differences, and based his conclusions on reported statistics and researcher interpretations. In another study, Falchikov and Goldfinch (2000) carried out a meta-analysis of 48 quantitative peer assessment studies that compared peer and teacher marks. Their study reported a mean correlation of .69 between peer and teacher marks, demonstrating that students are generally able to make reasonably accurate judgments.

Some authors have questioned the reliability and validity of studies such as those reviewed here, since most involved only a limited number of participants and compared student self-judgments with single teacher or peer judgments, thus lacking measurement reliability. Another shortcoming of the existing studies is that most had relied on single judgments by students and of teachers or peers. Furthermore, judgments by teachers or peers usually take place only towards the end of a task or course and are thus limited by recall (Falchikov, 2005; Ward, Gruppen, & Regehr, 2002). Van Daalen (1999), therefore, suggested that correlating self assessments with judgments of multiple assessors or with averaged teacher or peer

judgments would improve on reliability and validity of such self-assessment measures.

Most of the self-assessment studies in higher education focus on student assessing their capacity to acquire content knowledge, and of the accuracy of their self-predictions of performance when compared with actual achievement. However, less is known of students' ability to make judgments about their own learning *process*, viz., the act of self-monitoring their learning development, identifying strengths and weaknesses, and adapting learning in light of experience and feedback from teachers and peers. In the present studies, self-assessment takes on this latter interpretation. It refers to the reflection, evaluation, and appraisal by learners of their own competence and performance *in the course of their learning* (Paris & Paris, 2001). Instead of comparing self-assessment with performance on achievement tests, the present studies examine student self-assessment accuracy by comparing their self-judgments with other measures of the learning process (particularly teacher and peer judgments).

In Study 1 to be reported below, we conducted a longitudinal study to examine whether students can learn to self-assess given that they repeatedly have to evaluate themselves as the semester unfolds, and receive continuous feedback from teachers and peers on their performance. Such provision of frequent feedback to students is judged to be optimal for learning self-assessment (Butler & Winne, 1995). Given the almost continuous nature of teacher and peer judgments in this study, the problem of them being limited by recall was avoided (Falchikov, 2005; Ward et al., 2002). In order to maximize the reliability of students' self assessments, we did not rely on single judgments of a selected, small group of students, nor of teachers and peers. Instead, all first-year students ($N = 3588$) were involved in the first study, and multiple self-, peer and teacher judgments were averaged and used in our analyses. In assessing the accuracy of student self-assessment ability, their self-judgments were correlated with the judgments by multiple teachers and peers (see also Van Daalen (1999)).

It was hypothesized that students' self-assessment ability would improve with experience, as they progress through the course,

engaging in continuous self-assessment. It was also conjectured that academically more competent students were able to self-assess with greater accuracy, since such students are assumed to have better developed self-assessment skills. Orsmond, Merry, and Reiling (1997b) and Cassidy (2007) reported preliminary findings linking self-assessment skills with intellectual capacity: academically able students were indeed able to self-assess with greater accuracy.

While Study 1 focused primarily on self-assessment accuracy and whether it increases with experience, Study 2 was conducted to examine if relationships exist between students' beliefs about self-assessment and their self-assessment ability. We will elaborate on the latter issue in the Discussion section of Study 1. In summary, the studies to be reported were conducted to elucidate three questions: (1) How accurate is self-assessment of competence and performance in the course of learning? To answer this question, student self assessments were compared with multiple judgments of teachers and peers. (2) Does the accuracy of self-assessment improve over time? To that end, the accuracy of students' self assessments was studied throughout one semester. In this semester, the students involved made approximately 80 self assessments each. And, (3) are self assessments more accurate if students believe that this activity really contributes to their learning? To that end, a questionnaire was administered to students to elicit their beliefs about the effects of self-assessment on their learning.

2. STUDY 1

Study 1 was conducted to examine students' ability to self-assess by comparing their self assessments with the judgments by their teachers and peers. It also sought to investigate whether students' ability to self-assess improves with experience and through engaging in continuous self-assessment as they progress through a semester.

2.1 Method

2.1.1 Subjects

Participants were 3588 students in their first year of studies at a polytechnic in Singapore in the academic year 2007-2008. Of these students, 1843 (51 %) were females and 1745 (49%) were males, and their mean age was 18.23 years ($SD = 1.44$). The Grade Point Average (GPA) is calculated based on students' classroom performance grades as awarded by their tutors, and their grades on knowledge acquisition tests. The GPA values which range from "A" to "F" were first converted to scaled numerical values on a five-point scale. The mean GPA value of the participants at the end of the first semester of the academic year was 3.39 ($SD = 0.47$).

2.1.2 Educational Context

Problem-based learning. The polytechnic at which the research was carried out organizes its curriculum according principles of problem-based learning. Students work collaboratively in teams of four to five, with learning centred on problems relevant to their domain of study. They work each day on one problem. The problem is initially discussed in the morning, followed by ample study. At the end of the day, information gathered is shared and elaborated upon. No didactic teaching takes place nor is there any form of direct instruction. One tutor supervises the student teams in a larger classroom. His or her role is to facilitate student learning (Alwis, 2007). There are two semesters in an academic year, with each semester lasting 16 weeks. All the programs offered are three-year curricula.

Assessment in the curriculum. The daily assessment approach consists of four, independent elements: (1) a judgment by the tutor on how well students have performed during the day (2) a self-assessment, and (3) a peer assessment, and (4) a reflection journal to be written by each student. Students also need to take four knowledge acquisition tests per module, which are taken at different points during the semester. The duration of each test is 30 minutes

and it consists of answering at least three structured questions. The tests are conducted in a supervised environment, similar to an end-of-course examination. Students are tested on their ability to understand and apply what they have learnt.

2.1.3 Instrument

The self-assessment rating scale consists of 8 items inquiring about the quality of students' performance within their team, such as the level of cooperativeness and contribution of ideas. A Cronbach's alpha value of .90 gives evidence for the high internal consistency reliability of the self-assessment instrument. The peer assessment rating scale consists of 4 items inquiring about the cooperativeness and quality of contributions of peers within the team. The peer assessment instrument has high internal consistency reliability, given its Cronbach's alpha of .93. In examining the inter-rater agreement by correlating the scores awarded to students by different peers, we computed the intraclass correlations based on students' peer assessment scores. Intraclass correlations of .97 and .95 for the first and the second semester respectively were obtained. The values of Cronbach's alpha were computed based on student responses on the items of the self- and peer assessment instruments in semester one of the 2007-2008 academic year. Students are asked to respond to these items on a Likert five-point scale ranging from "strongly agree", "disagree", "neutral" and "agree" to "strongly agree". The items for the self- and peer assessment are contained in Appendix A. On a particular day, each student assesses and is in turn assessed by his peers within the team.

The tutor judgment consists mainly of tutors' observations of students' processes of daily learning. The observations by the tutors include students' self-directedness, level of participation inclusive of teamwork; students' ability to reason, justify and defend opinions and ideas formulated in respond to problems, as well as their problem solving skills. Tutors will then award grades ranging from "A" to "F", which are derived based on what they observe and the impression they have on each student during the duration of time they had with him/her. Tutors also take into consideration students'

individual reflection journals (short essays which document students' reflections on daily learning) and their self and peer assessments when awarding grades. Furthermore, tutors will provide feedback to students on their learning outcomes and processes of daily learning. The generalizability of judgments made by different tutors is high, with an average generalizability coefficient of .84 (Chua & Schmidt, 2007).

2.1.4 Procedure

Data used in the analyses was (students' self and peer assessments and their tutor grades) collected in the first semester of the academic year 2007-2008. The tutor grades were first converted to scaled numerical values on a five-point scale. To analyze the data collected in a meaningful fashion, the raw data collected was clustered and averaged over time intervals of four weeks each. The averaged values of the grouped self, peer and tutor assessments were used for the analyses.

2.1.5 Analysis

Descriptive statistics (means and standard deviations) of students' self- and peer assessment scores and tutor grades were computed. Correlational analyses were performed to examine the inter-relationships between students' self, peer and tutor assessments. Correlations were also computed for the scores of selected groups of students: the low-achieving and the high-achieving students. The low- and high-achieving students were selected based on their GPA values. Low achievers represented the bottom 10% of the first-year student cohort, whereas the high achievers represented the top 10% of all first year students.

2.2 Results

The output of correlation analyses for students' self- and peer assessment scores and their grades by the tutors are contained in

Table 1. The correlation coefficients (r) were computed based on the overall ungrouped data of the scores for the entire semester.

Table 1. Overall correlations between students' self, peer and tutor assessments

	Self-assessment scores	Peer assessment scores
Peer assessment scores	.31**	
Tutor grades	.23**	-.03

Note. Correlations were computed based on ungrouped data for the entire semester.

** $p < .01$, 2-tailed

The mean overall correlations between students' self- and peer assessment scores, and that of their self-assessment scores and tutor grades are $r = .31$ and $.23$ respectively. Correlations coefficients between scores of self and peer assessments ranged from $.17$ to $.40$, and those of self-assessment scores and tutor grades ranged from $-.07$ to $.31$. These r -values suggest that a moderate inter-relationship exists between the judgments by students of their own learning process and those by their peers; and a weak inter-relationship exists between students' self-judgments and their tutor judgments. By and large, the correlations indicate moderate to weak accuracy of student self-assessment ability as witnessed by how they were judged by their peers and tutors.

Table 2. Correlations of self-, peer and tutor assessments of low-achieving and high-achieving students

	Self-assessment scores	
	Low achievers ^a	High achievers ^b
Peer assessment scores	.23**	.41**
Tutor grades	.01	.29**

Note. Low achievers refer to the bottom 10% of all first-year students with GPA ranging from 2.01 to 2.89. High achievers refer to the top 10% of students with GPA ranging from 3.93 to 4.77.

^a $n = 359$. ^b $n = 368$

** $p < .01$, 2-tailed

The correlations between students' self- and peer assessment scores are higher for high-achieving students as compared to low-achieving students ($r = .41$ and $.23$ respectively.). Similarly, the correlation between students' self-assessment scores and tutor grades for high-achieving students is higher for high achievers as compared to low achievers ($r = .29$ and $.01$ respectively.) A method that compares correlations drawn from different samples as described by Hays (1988) was used to test for significant differences between them (p.591). The difference between these r -values was in both cases statistically significant ($p < .001$).

Descriptive statistics for students' self- and peer assessment scores and tutor grades are given in Table 3.

Table 3. Descriptive statistics of students' self-, peer and tutor assessments as a function of time (per 4 weeks)

	Self-assessment scores		Peer assessment scores		Tutor grades	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Weeks 1-4	3.91	.41	4.18	.39	3.70	.48
Weeks 5-8	3.96	.43	4.20	.40	3.62	.56
Weeks 9-12	3.97	.43	4.21	.41	3.62	.61
Weeks 13-16	3.98	.43	4.22	.42	3.41	.70

Note. All scores were calculated based on a 5-point scale and were averaged over fixed time intervals of 4 weeks. *SD* = standard deviation.

The mean values of peer assessment scores were the highest, followed by students' self-assessment scores and their grades by the tutors. The plots of self- and peer assessment scores with time (per 4 weeks) depict a gradual increase of scores from weeks 1 to 16. On the contrary, the mean values of the tutors' grades decrease from week 9 onwards (Figure 1). By and large, all the scores remain relatively stable throughout the semester as represented by the nearly horizontal plots of mean scores with time. The mean differences between the grouped data (self- and peer assessment scores and tutor grades) for the periods from weeks 1 to 4 and that from weeks 5 to 8 were tested for significant differences using paired-samples t tests. The outcomes of the analyses demonstrate that the differences between the grouped data are statistically

significant (self-assessment: $t(3587) = -8.11, p < .01$; peer assessment: $t(3587) = -4.74, p < .01$; tutor grades: $t(3587) = 12.15, p < .01$, with degrees of freedom in parentheses). Of course, given the sample size of the current study, extremely small and insignificant differences can be found to be statistically significant.

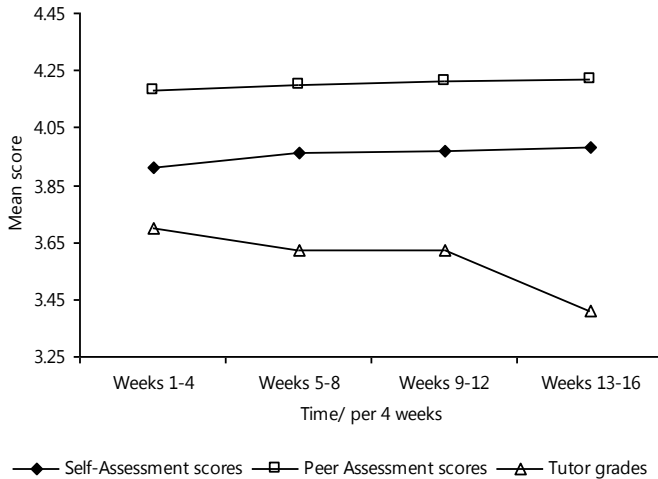


Figure 1. Plots of students' mean self- and peer assessment scores and tutor grades as a function of time

Table 4 contains the results of the correlational analyses between students' self- and peer assessment scores and tutor grades as they progressed through the semester. A gradual, decreasing trend in the linear relation between the self- and peer assessment scores is observed, with moderate r -values ranging from .37 to .29. A similar change pattern is noted for students' self-assessment scores and their tutor judgments, with low r -values ranging from .27 to .15. The correlations between students' self-assessment scores and tutor grades are comparatively lower compared to students' self-judgments and the judgments by their peers.

Table 4. Correlations between students' self-, peer and tutor assessments over fixed time intervals as students progressed through the first semester

	Self-assessment scores			
	Weeks 1-4	Weeks 5-8	Weeks 9-12	Weeks 13-16
Peer assessment scores	.37**	.33**	.31**	.29**
Tutor grades	.27**	.21**	.18**	.15**

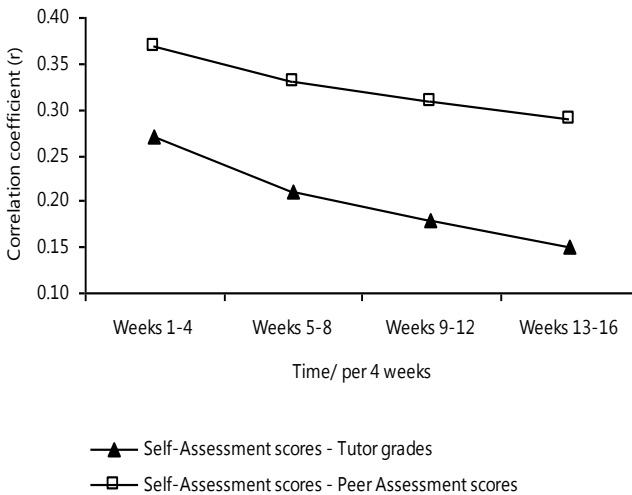


Figure 2. Plots of the correlations between students' self and peer assessment scores, and their self-assessment scores with tutor grades as a function of time

The correlations contained in Table 4 are graphically represented by Figure 2. As observed, the r -values seem to decrease throughout the semester. A method that compares correlations drawn from the same sample as described by Cohen and Cohen (1983) was used to test for significant differences between them (p.57). Results of the analysis reveal that the differences in the correlations between self- and peer assessment scores computed for different time intervals were not

statistically significant. Similar findings were obtained for the correlations between self-assessment scores and tutor grades.

2.3 Discussion

Study 1 was conducted to examine the accuracy of student self-assessment ability as compared to how they were judged by their peers and tutors. To that end, students' self assessments were compared with multiple judgments by peers and tutors. The overall correlations of students' self- and peer assessment scores and tutor grades for the entire semester suggest weak to moderate accuracy of student self-assessment ability, as witnessed by how they were judged by their peers and tutors. The overall correlations between scores also suggest an ability effect, where students judged as being more competent academically (according to GPA values) were able to self-assess with greater accuracy as compared to their less competent peers. Furthermore, we were interested to investigate if self-assessment ability of students improves over time. The mean plots of students' self-, peer and tutor assessment scores for various time intervals reflect that students underestimated their own performance in comparison to how they were assessed by their peers. Students however, overestimated their own performance as compared to the grades given by their tutors (Figure 1).

There are at least three possible explanations for these findings. First, there is the possibility that students are generally poor assessors. They simply are not able to perform the task accurately, for instance because they have insufficient access to their own learning process. However, a study by Sullivan and Hall (1997), for example, has demonstrated that students can be competent self-assessors; they reported a correlation of $r = .72$ between students' self assessments and the marks awarded by their teachers. So, a general dismissal of the idea that students are competent self-assessors may be premature.

A second possibility is that students in general may be fairly good at self-assessment, but the students in this particular study are somehow lacking the experience. Students who took part in the current study could be described as 'inexperienced' to some extent,

because they were first-year students in their first semester of study in higher education. Falchikov and Boud (1989) suggest that experienced students, i.e., those in their later years of studies, were able to provide more accurate self assessments than those students in introductory programs, and Gibbs (1995) contends that students entering higher education indeed self-assess with weaker accuracy as compared to more experienced students in higher years. In another study, Cassidy (2007) however did demonstrate that first-year students in their first semester of higher education already have the capacity to allow for accurate self-assessments. In addition, our participants may have been new to higher education, although they already had more than ten years of education behind them. Nevertheless, one cannot exclude the possibility that the beginning of a new study is not the best moment to look at self-assessment accuracy and that the duration of measurement (16 weeks) was too short to observe significant increases in student self-assessment ability. To see if a longer period would give rise to more meaningful findings, we examined post-hoc the data of students' self, peer and tutor assessments for the *second* semester. Again, weak overall correlations between self-and peer assessment scores, and self-assessment scores and tutor grades were obtained ($r = .26$ and $.21$ respectively). Test of differences between these correlations and those obtained from the first semester revealed no significant differences. In examining if student self-assessment ability improves over time, correlations between scores over fixed time intervals (per 4 weeks) were also computed. The correlations obtained suggest weak inter-relationships between students' self assessments and how they were judged by their peers (r ranging from $.21$ to $.26$) and tutors (r ranging from $.16$ to $.22$). The differences between these correlations and those obtained from the first semester were not significant. The findings obtained based on the second semester data suggest that students' ability to self-assess not only does not improve, but, in fact, appears to become worse over time.

A third possible explanation for our findings is, students may overall be competent self-assessors but correlations are weak to moderate because tutors and peers judge upon only a portion of the learning behaviours of the students involved (i.e. all the instruments

used in this study showed only partial overlap). To test for this possibility, we examined the items of the self- and peer assessment instruments more closely. These items seem to have two aspects of the learning behaviours of students in common, namely, cooperativeness and what we would call 'idea contribution'. What if we would compute correlations based only on these two features of student learning? To that end, the first, second, third, and eighth items from the self-assessment instrument were grouped to form the learning behaviour of cooperativeness. Items 4-7 on the self-assessment instrument were grouped to form the idea contribution learning behaviour. Students' self-assessment scores on the identified learning behaviours were then correlated with peer assessment scores (cooperativeness with item one of peer-assessment instrument and idea contribution with item four). Moderate inter-relationships emerged between self- and peer assessment scores (cooperativeness: $r = .38$; idea contribution: $r = .30$). Differences between these correlations with those reported in Table 1 are not statistically significant.

To deal with the issue of a possible mismatch between the instruments more conclusively, we conducted a follow-up study. In this preliminary study among 400 randomly selected students, we compared self-, peer and tutor assessment instruments, which were evaluating identical aspects of student performance. Again, weak correlations emerged between students' self and peer assessments ($r = .19$) and that of self- and tutor assessments ($r = .25$). Differences between these correlations and those reported in Study 1 were not statistically significant, suggesting that the use of instruments for self-, peer and tutor assessment intended to measure similar aspects of performance, does not improve student self-assessment accuracy significantly.

These deliberations lead us to the conclusion that, generally, students are fairly poor in judging their own learning process accurately, and that that this skill cannot easily be learnt. This conclusion does not necessarily apply to all students. We found that students' ability to self-assess is closely related to intellectual capability. It is cogent to argue that students judged as being more competent academically are inclined to self-assess more accurately

given that they are better at self-monitoring, judging their own performance and processes of learning, and at identifying their own learning strengths and weaknesses. The correlations reported between students' self, peer and tutor assessments in the current study suggest that high-achieving students have better developed self-assessment skill, making them more competent self-assessors as compared to low-achieving students (see also Boud & Falchikov (1989); Orsmond, Merry, & Reiling (1997b)). Falchikov and Boud (1989) have also reported that more academically competent students were able to self-assess with greater accuracy. Such a finding linking self-assessment skill with intellectual capability is therefore not unexpected, since many authors have associated self-assessment with self-regulated learning, metacognitive and self-reflective thinking (Mok et al., 2006; Paris & Paris, 2001).

A final possible explanation for the fairly poor accuracy of self-assessment not yet discussed here is that some students simply do not take the self-assessment activity seriously while others perhaps do, leading to overall weak to moderate accuracy of these judgments. The way in which students respond to self-assessment may be due to how they perceive it and which in turn will determine the way they tackle their own learning (Segers & Dochy, 2001). It is, therefore, suggested here that students who do not believe that self-assessment contributes towards improving learning, for instance because they regard assessing themselves as a 'mechanical, meaningless task' (Lew and Schmidt 2006; Maguire, Evans & Dyas, 2001), may be less accurate than those who believe that self-assessment does contribute. Therefore, the relationships between students' beliefs about self-assessment and the accuracy of their self assessments were examined in greater detail in the subsequent study.

3. STUDY 2

Study 2 was conducted to test hypotheses predicting interactions between students' beliefs about self-assessment and their ability to self-assess as compared to the judgments made by peers and tutors. It is conjectured here that a relationship exists between students'

beliefs about the effects of self-assessment on their learning and their ability to self-assess accurately. First, students who believe that their learning improves through self-assessment are hypothesized to act accordingly (treat it more seriously), and learn more from it. Through the process of reflecting on their learning, these students become better aware of their strengths and weaknesses, enabling them to take steps to further improve. As a result, their performance improves. Conversely, students who do not believe that self-assessment contributes to their learning are hypothesized to tend to take it less seriously and hence, they do not learn from the process of assessing themselves. Their performance is not expected to improve.

3.1 Method

3.1.1 Subjects

Participants were 936 first-year students in the academic year 2007-2008. Of these 936 students, 477 (51%) were female students and 459 (49%) were male students. The mean age of the participants was 18.33 years ($SD = 1.54$), while their mean GPA value was 3.42 ($SD = .40$). The participants were representative of the entire cohort of 3588 first-year students. They were then clustered into sub-groups based on their responses on the self-assessment section of a questionnaire used in the study.

3.1.2 Educational context

Study 2 was conducted in the same institution as Study 1.

3.1.3 Instrument

Students responded to a 25-item questionnaire containing statements inquiring about their beliefs about self-assessment which was developed and validated in an earlier study (Lew and Schmidt 2007b). The questionnaire consists of seven belief categories derived from past research studies on the use of self-assessment tools and reflection journals in higher education. For the purpose of this study,

only the self-assessment section of the questionnaire will be discussed.

The self-assessment section consisted of 10 items and three underlying constructs (see Appendix D). A study by Mok et al. (2006) revealed that students become better aware of their learning and thinking process through engaging in self-assessment, thereby enabling them to take steps to improve on their learning deficiencies. Items such as 'The self-assessment helps me to assess my strengths and weaknesses accurately' served to measure the construct of 'usefulness of self-assessment in appraising students' learning.' Lew and Schmidt (2006) reported that students sometimes seem to use the self-assessment as an impression management tool. The impression management construct was measured by items such as 'The self-assessment is mainly useful in managing the tutor's impression of my performance.' They also highlighted the issue that students lacked conscientiousness and were not seriously contemplating the task of assessing their own learning. This belief category was measured by items such as 'I do the self-assessment without thinking how the statements are related to my performance during the day.'

3.1.4 Procedure

The questionnaire was administered online to the participants in the tenth week of the semester. The questionnaire's instruction stated that there were no right or wrong answers to the items, all answers were correct as long as they reflected students' opinions. No information was given regarding the constructs underlying the questionnaire.

3.1.5 Analysis

In examining if relationships existed between students' beliefs about the usefulness of self-assessment and the accuracy of their self-assessment ability, correlational analyses were performed. Students' responses on those items measuring the constructs of the usefulness of self-assessment as tools for learning and impression management

were related to the accuracy of their self- assessment expressed as correlations with peer assessment scores and tutor grades.

In performing the analyses, selected groups of students were identified by comparing their mean responses on the items measuring the constructs on the usefulness of self-assessment in appraising learning and in managing tutors' impressions of them. The students who hold strong beliefs are those with top 25% of mean responses whilst those students who hold weak beliefs have mean responses in the bottom 25%.

3.2 Results

Table 5 contains the correlation coefficients between self- and peer assessment scores and the tutors' grades of two groups of students: those students who hold strong beliefs that they become better aware of their learning and thinking processes through engaging in self-assessment, thereby enabling them to take steps to improve on their learning deficiencies and, those students who hold weak beliefs that self-assessment is useful in aiding their learning. Moderate inter-relationships exist between students' self assessments and the judgments of students by their peers and tutors (r ranging from .22 to .39) for both student groups. Tests of differences between the correlations for both student groups reveal that the differences were not statistically significant.

Table 5. Study of students' beliefs about self-assessment as a learning tool and their ability to self-assess accurately

	Self-assessment scores	
	Useful for learning	Not useful for learning
Peer assessment scores	.39**	.35**
Tutor grades	.33**	.22**

Note. ^an = 404. ^bn = 404

** $p < .01$, 2-tailed

Correlations between students' beliefs about self-assessment as an impression management tool and their scores on the various measures of performance and learning are contained in Table 6.

Table 6. Study of students' beliefs about self-assessment as an impression management tool and their ability to self-assess accurately

	Self-assessment scores	
	Useful for impression management	Not useful for impression management
Peer assessment scores	.43**	.32**
Tutor grades	.19**	.21**

Note. ^cn = 366. ^dn = 371

** $p < .01$, 2-tailed

Outcomes of the analyses demonstrate that moderate relations exist between the self- and peer assessment scores for those students who believe that the self-assessment is useful as an impression management tool ($r = .43$) as compared to those who believe otherwise ($r = .32$). The results also show that low relations exist between self-assessment scores and tutor grades for those students who hold strong beliefs that self-assessment is useful as an impression management tool and the scores of those students who hold weak beliefs ($r = .19$ and $.21$ respectively). The differences between these correlations were however not statistically different.

In an attempt to select groups of students with even more strongly positive or negative beliefs about the utility of self-assessment, we also selected groups that were either high or low in *both* categories. One group consist of those students who believe strongly in the usefulness of self-assessment as a learning tool, but hold weak beliefs about its usefulness as an impression management tool. On the other hand, those students who believe strongly in the usefulness of self-assessment as an impression management tool and hold weak beliefs about its usefulness as a learning tool formed another group. The results are displayed in Table 7.

Table 7. Study of students' beliefs about self-assessment as both learning and impression management tools and their ability to self-assess accurately.

	Self-assessment scores	
	Useful for learning but not for impression management	Useful for impression management but not for learning
Peer assessment scores	.30*	.32*
Tutor grades	.42**	.11

Note. ^en = 44. ^fn = 47

** $p < .01$, 2-tailed, * $p < .05$, 2-tailed

A moderate inter-relationship exists between self-assessment scores and tutor grades for students who believed in the usefulness of the self-assessment as a learning but not as an impression management tool ($r = .42$). By contrast, no significant relation between these scores for students who believed in the usefulness of the self-assessment for impression management and not for learning is reported. Testing for differences among the correlations between self-assessment scores and tutors' grades for both student groups indicated that these were not statistically significant.

The correlations between students' self, peer and tutor assessments over fixed time intervals for selected groups of students are contained in Table 8. Testing for differences among the correlations for the groups indicated that these were also not statistically significant.

Table 8. Correlations between students' self, peer and tutor assessments over fixed time intervals for selected groups of students

Student group	Self-assessment scores			
	Weeks 1-4	Weeks 5-8	Weeks 9-12	Weeks 13-16
Useful for learning				
Peer assessment scores	.43**	.39**	.38**	.33**
Tutor grades	.31**	.26**	.30**	.25**
Not useful for learning				
Peer assessment scores	.40**	.38**	.31**	.32**
Tutor grades	.25**	.16**	.19**	.14**
Useful for impression management				
Peer assessment scores	.48**	.44**	.40**	.38**
Tutor grades	.24**	.12*	.19**	.16**
Not useful for impression management				
Peer assessment scores	.38**	.35**	.29**	.28**
Tutor grades	.27**	.16**	.18**	.13*
Useful for learning but not for impression management				
Peer assessment scores	.29	.38*	.30	.26
Tutor grades	.38*	.30	.28	.28
Useful for impression management but not for learning				
Peer assessment scores	.40**	.30*	.34*	.43**
Tutor grades	.40**	.01	.09	.06

** $p < .01$, 2-tailed, * $p < .05$, 2-tailed

3.3 Discussion

In Study 2, we examined if a relationship exists between students' beliefs about the effects of self-assessment on their learning and the accuracy of their self assessments. Selected groups of students were identified by comparing their mean responses on a questionnaire

measuring the constructs on the usefulness of self-assessment as a learning tool and, alternatively, an impression management tool.

Comparisons among the correlations of students' self assessments with the judgments by peers and tutors for all student groups revealed none to be statistically significant. These findings suggest that there are no inter-relationships between students' beliefs about the usefulness of self-assessment and their self-assessment ability. Furthermore, students do not show improvements in their self-assessment ability over time (Table 8). Thus, our findings suggest that the accuracy of self-assessment is no different for students who hold strong beliefs and for those who hold weak beliefs about the effects of self-assessment on their learning. So, whatever students believe about the effects of self-assessment on their learning, no effects can be observed on their self-assessment accuracy.

4. GENERAL DISCUSSION

The present research was conducted to study the self-assessment accuracy of students of their own learning *process*, the changes in accuracy over time, and to examine whether inter-relationships exist between students' beliefs about self-assessment and their self-assessment accuracy. In the first study, students' self assessments were compared with other measures of their performance such as judgments by their peers and tutors. Overall correlations between students' self- and peer assessment scores, and their self-assessment scores and tutor grades indicate weak to moderate accuracy of student self-assessment ability. The findings also indicate that students judged as more competent academically were able to self-assess with higher accuracy as compared to their less competent peers. Such a finding linking self-assessment skill with intellectual capacity is mirrored in several other studies (Falchikov & Boud, 1989; Orsmond et al., 1997b). Furthermore, comparing the accuracy of student self-assessment averaged over four consecutive periods indicates that it does not improve over time.

In the second study, students' who held either strong or weak beliefs about the usefulness of self-assessment as a learning tool and/or an impression management tool were identified, and their

self-assessments compared with judgments by their peers and tutors. The results suggest that there is not significant association between student beliefs about the utility of self-assessment and the accuracy of their self-assessments. There appears to be no differentiation in the accuracy of self-assessment ability between those students who hold strong beliefs and those who hold weak beliefs. These findings seem to suggest that students' beliefs about the use of self-assessment are not relevant to the development of self-assessment skills; however, more research is necessary here.

Taken together, our findings indicate that students on average do possess accurate self-assessment skills only to a limited extent. In addition, our studies provide evidence that self-assessment is not learned through extended experience and regular feedback. Our findings are to a large extent, in agreement with what Eva and others (2004) report about student self-assessment accuracy decreasing with increased seniority in the program (see also Fitzgerald, White, and Gruppen (2003)). In their work, Eva et al. (2004) demonstrated that student self-assessment skill does not improve after more than two years of general experience with a self-assessment environment, and upon receiving regular feedback from their teachers about previous performances on examinations which test their understanding of broad medical knowledge. Nonetheless, our findings are not agreement with what Dochy et al. (1999) reported namely, that student self-assessment accuracy does improve over time and with practice.

Contrary to most self-assessment studies with limitations such as small sample size, non-continuous student self assessments or infrequent feedback given by teachers, the present studies have sought ways in arriving at more stable and reliable measurements. We did not rely on single self-judgments of students, nor of teachers and peers, and adopted the recommendations by Van Daalen (1999) to enhance the reliability of the findings. Furthermore, in this context, students received continuous feedback on their performance from peers and teachers. Such provision of regular and timely feedback must have created optimal conditions for learning self-assessment, given that students receive explicit cues and suggestions from their peers and teachers on how they can further improve on their

learning. However, despite all efforts taken to ensure reliability of measurement, and to optimize feedback for learning, we failed to find any sizable effect.

4.1 Limitation

A shortcoming of the present studies is the partial overlap of the instruments used: self-assessment, peer assessment and tutor judgment, which may have produced, in part, the weak to moderate correlations between students' self-, peer and tutor assessment scores. Although we have tested the hypothesis that partial overlap between measures may have been a cause of weak correlations, and had to reject it, a study employing identical instruments for self- and peer assessment should certainly be conducted to verify our findings.

4.2 Future Research

Two other issues present themselves for further research based on the findings from the present studies. First, given the range of students' aptitude and ability to cope with, and respond to, the task of assessing their own learning, the focus on individual students and their strengths and weaknesses should constitute the next stage of research in better understanding the nature and operation of self-assessment in higher education. The gathering of detailed empirical evidence which may cast light on those characteristics and factors which could account for individual differences in student self-assessment skill is one key area for further research.

Second, further research should investigate if student self-assessment skills can be improved through formal training in self-assessment. Feedback alone, as our study has demonstrated, is clearly not enough to affect change. Through a more structured and closely guided process, students may become better aware of, and value their existing capability for, self-assessment, and its potential for development and application. If students have better developed self-assessment skills, it is likely that they will involve themselves in more effective learning and will thus become better metacognitive

and self-reflective learners capable of critical evaluation of their own performance, a skill so highly valued in professional practice.

Chapter 5- Writing to learn and learning to write: Does reflection journal writing improve student learning?

ABSTRACT

The purpose of the present study was to evaluate whether there is evidence of reflection in student-written journals, and to investigate whether students show improvements in their reflective and writing skills through journal keeping. To that end, the reflection journals of 3460 first-year students enrolled in a post-secondary institution was studied by means of an automated coding procedure using software. Data used in the analyses were students' journals for an entire week, collected once at the beginning, and again, at the end of an academic year. Outcomes of the content analyses suggest that there is evidence of reflection in students' journals, and they reflected on three general categories related to their learning: critical review of past learning experiences, learning strategies and summaries of what was learnt. Furthermore, the findings also indicate that students show improvements in their reflective skills as they progressed through the academic year. In examining changes in writing skills throughout the year, outcomes of paired-sampled *t* tests suggest that students wrote simpler journals with higher readability and more spelling and grammatical mistakes. In addition, students' written expression of ideas in a more coherent manner did not improve even after engaging in journal writing for almost a year.

Keywords: *Reflection journals, self-reflection, metacognition, text analysis, writing skills*

1. INTRODUCTION

The incorporation of reflection journals as learning and assessment tools into programmes of study within higher education arises from the recognition of the possible positive roles that reflection may play in fostering students' self-reflection, critical thinking, creative writing abilities, and in the demonstrable development of professional values or skills (Hubbs & Brand, 2005; Morrison, 1996). More recently, there has been a growing interest in reflection journals to be used as part of a reflexive metacognitive strategy. Reflection journal writing is believed to enable students to critically review *processes* of their own learning and behaviours, and to understand their ability to transform their own learning strategies (Gleaves, Walker, & Grey, 2008a, 2008b).

Reflection journals are variously referred to "reflective journals" (e.g. Chirema, 2007), "reflective learning journals" (e.g. Thorpe, 2004) or "learning journals" (e.g. Creme, 2005; Langer, 2002). Although they are used in a variety of courses, reflection journals are essentially written records that students create as they think about various concepts learnt, about critical incidents involving their learning, or about interactions between students and teachers, over a period of time for the purpose of gaining insights into their own learning (Thorpe, 2004). Several authors (e.g. Dart et al., 1998; Haigh, 2001; Voss, 1988) have emphasized that reflection journals, by focusing on the *processes* (e.g. self-reflection and learning strategies) rather than the *products* (e.g. summaries of knowledge acquired) of learning, may enable students to improve on their reflective capacity and skills.

Why is reflection deemed as important? Increasingly, many institutions of higher education have introduced reflective practices into their courses as the ability to reflect on one's knowledge and experience is valued by many as a means of dealing with the complexities, challenges, and uncertainties inherent in professional life (Langer, 2002; Moon, 1999a; Thorpe, 2004). There is a widely-documented view that self-reflection enhances professional practice, since the learner is involved in processes which explore experience as a means of deepening understanding (Boud, Keogh, & Walker, 1985; Moon, 1999a). These processes include "looking back on experiences, decisions and actions; recognizing values and beliefs underlying these actions and decisions; considering the consequences and

implications of beliefs and actions; exploring possible alternatives; and reconsidering former views" (Sumsion & Fleet, 1996, p.121). Sumsion and Fleet further contend that since processes as these can lead to informed and thoughtful deliberation on one's beliefs and actions, they are expected to assist learners in becoming reflective practitioners. Boenink and others (2004) even go on to emphasize that in order for the development of a balanced professional identity, self-reflection is a necessary prerequisite.

The use of reflection journals as a learning tool therefore highlights the role of self-reflection in learning. Self-reflection (or simply, reflection) has received numerous definitions from different sources in the literature. Depending on the emphasis on theory or practice, literature definitions vary from philosophical articulations as in John Dewey (1991), formulations in theoretical frameworks, such as the "reflection-in action" and "reflection-on-action" constructs developed by Schön (1983), to the use of reflection in the experiential learning cycle by Kolb (1984). What further complicates the picture of self-reflection is the range of different purposes or outcomes that the activity of reflecting seems to fulfil. Besides seeking to develop metacognition in students, other purposes of journal keeping include: to critically review the behaviours (e.g. strengths and weaknesses; learning styles and strategies) and learning of self and others, or the products of learning; to set or track learning goals; to explore connections between knowledge that was learnt and students' own ideas about them; and, to improve writing skills (Langer, 2002; Moon, 1999a).

The definitions and purposes of self-reflection, though heterogeneous, are united in their advocacy for improving student learning. In his work, Zimmerman (2000) argues that self-reflection plays a critical role in achieving self-regulation in learning. Several other authors (e.g., Paris & Cunningham, 1996; Paris & Paris, 2001) are also in favour of Zimmerman's argument. They support the premise that processes of self-regulated learning enable the learner to monitor, direct and regulate his actions towards goals of information acquisition, expanding expertise and self-improvement. Self-regulated learning can be taught indirectly with classroom activities or by using tools to evoke self-reflection, cognitive and metacognitive understanding. The different purposes and outcomes

of self-reflection closely match many of the purposes for reflection journal keeping. Therefore, journal writing represents a formal tool to encourage reflection and metacognition (Langer, 2002). It is hoped that through reflecting and writing about new information or ideas, learners can better understand and remember them, and that the articulation of connections between new information, ideas, prior or existing knowledge also deepens learning (O'Rourke, 1998).

A particular emphasis on cognition and metacognition is evident in the research on factors influencing effective classroom learning (Marton & Säljö, 1984). Several studies have shown that the use of cognitive strategies enhance learning across a variety of domains (e.g. McCombs & Whistler, 1989; Weinstein & Mayer, 1986). Weinstein and Mayer have identified three groups of learning strategies that enhance a learner's cognition: rehearsal, organization and elaboration. *Rehearsal* involves the learner in repetition of to-be-learned information in a form relatively unchanged from the form in which it was given. Oral repetition, copying and making selective verbatim notes are some examples of rehearsal strategies. *Organization* involves the learner in rearrangement of to-be-learned information in a way that makes meaning more meaningful. Examples of organizational strategies include categorizing and constructing networks. *Elaboration* involves the learner in integration of presented information with prior knowledge. Examples of elaboration strategies include paraphrasing and summarizing. Weinstein and Mayer have shown in their research that all three types of strategies described enhance the acquisition and retention of information. Other strategies that learners use to focus attention, and to establish and maintain motivation are also prototypical of research on effective classroom learning.

Journal keeping has also been positively associated with enhancing student metacognition. In his work, Vockell (2004) describes metacognitive skills as the learners' automatic awareness of their own knowledge and their ability to understand, control and manipulate their own cognitive processes (see also Flavell (1979)). In reviewing the literature in the past century on teaching and learning, the American Psychological Association (1997) highlighted metacognition as one of the more important factor in contributing towards effective learning. The review suggests that as students'

metacognitive abilities develop, so does their ability for self-reflection and self-regulation of learning, thus leading to improvements in academic performance. This is illustrated in the work by McCrindle and Christensen (1995), whose study explores the impact of journal writing on cognitive and metacognitive processes, and academic performances of forty undergraduates in a first-year biology course. Students were randomly assigned to a learning journal (experimental) group or scientific report (control) group. Their findings demonstrate that students in the experimental group used more cognitive and metacognitive strategies during a learning task as compared to those in the control group. Students who kept learning journals also showed more sophisticated conceptions of learning, greater awareness of cognitive strategies, and demonstrated the construction of more complex and related knowledge structures when learning from text. Furthermore, they also performed significantly better on the final examination for the course.

The literature offers evidence that students, regardless of their domains of study, show improvements in their learning, *viz.*, students becoming better in self-assessment, through journal keeping. For instance, Selfe, Petersen, and Nahrgang (1986) investigated the use of journals in a college-level mathematics course. Their study showed that while journals did not necessarily assist students with earning high grades on achievement tests, journals did assist students in developing abstract thinking thereby enabling them to better conceptualize the meaning of technical definitions. Furthermore, students appeared to develop better strategies in problem solving through writing as compared to mere memorizing of calculations. An implication which arises from the study by Selfe and his co-workers is that the positive effect of journal writing on student learning is not necessarily measured by achievement test grades. The findings by Selfe and others were mirrored in the work by Moon (1999b), in which she summarized a number of studies which examined the effects of journal writing on student academic achievement across a variety of disciplines. In all, her work demonstrated the influence of journal keeping on student academic performance was subtle and did not seem to assist students with achieving better achievement test grades. Instead, journal keeping seems to facilitate student

learning in a number of ways, among them synthesizing new knowledge about a domain subject with their prior knowledge and learning, and recording of useful strategies in solving problems. In addition, students also showed improvements in their writing skills, for instance, they were able to develop personal conceptual definitions that were more understandable than technical definitions (see also Herrero (2007)).

By contrast, other researchers are less optimistic about the effects of journal writing on student learning. Woodward (1998) describes a study in which all students in an undergraduate teaching course had to keep journals in all their subjects. Students were asked to be reflective about their learning and practicalities of teaching through theory into practice. Close examination of students' journal responses revealed that they were far from reflective and were merely diary entries describing an event or activity. The findings by Woodward are mirrored in the study by Bain et al. (1999), who examined the effectiveness of using journal keeping enhancing the reflectivity of student teachers during field experience placements. Bain and others reported that students' journal responses were mainly descriptive; what had happened and what may be done differently were documented, but deeper questions of how and why were left unasked. Others problems in journal writing reported are the use of reflection journals as instruments for attacking fellow students or writing only what the teachers would like to read (Lew & Schmidt, 2006).

Besides the potential effects of journal writing on student learning suggested by the papers reviewed above, there is a body of literature reporting empirical studies on the assessment of reflection in learning. In light of this type of studies, research typically looks into assessing the level of reflective thinking from students' journals by means of coding schemes. For instance, Kember et al. (1999) adapted John Mezirow's (1991) categorization scheme for estimating the quality of reflective thinking in students' written journals. However, in the study by Kember and his co-workers, their coding scheme was developed based on the journal responses of only three first-year undergraduate students in a health-care course, thus there were issues with the reliability and generalizability of their findings. In another study, Wong and colleagues (1995) reported the use of a

framework to allocate adult nursing students to three categories of non-reflector, reflector, and critical reflector based on their journal responses. The authors described the coding scheme as a mirror of the conceptual frameworks developed by Boud et al. (1985) and J Mezirow & Associates (1990). Though the method was well documented, the authors cautioned that identifying textual elements within student journals and allocating them to finer levels of reflection within Boud's model was a difficult process and felt that the levels of reliability they could achieve were not suitable for recommending for further use by others.

Some authors have questioned the reliability and validity of studies such as those reviewed here, since most involved only a limited number of participants who engaged in journal writing only on a few occasions throughout a course. In order to ensure that coding of student journals is done in a standardized manner, a detailed set of coding instructions must first be created to guide the work. This makes the task time-consuming and expensive. Even then, there may be disagreements among coders on how to categorize specific responses, reducing the reliability of the resulting data. Another shortcoming of the existing studies is therefore that of inter-coder reliability, since coding is performed manually and thus dependent on a high degree of interpretation. Furthermore, no instances of authors adopting coding procedures by others can be located in the literature, suggesting the absence of a widely accepted coding scheme that can be used to assess reflection in student journals. To add on, the coding procedures described in many studies lacked details on how they were carried out, or were too complicated for use in analyzing large number of journal entries.

The present study seeks to determine whether there is evidence of reflective activities in students' journal responses. We were interested to find out whether students' journals, in agreement with what the literature suggests, focus on critical reviews of their own and/or that of their peers' processes of learning and behaviours, and understanding of learning strategies used to enhance their metacognition. Furthermore, we were also interested to find out to what extent students use reflection journals to summarize the content of what they have learnt. This is considered by some as ineffective in enhancing students' reflective abilities, as it is thought

to hinder the objective of developing critical thinking and metacognitive skills (Langer, 2002; Voss, 1988). But to what extent does it actually emerge in these responses?

A second goal of the study was to investigate whether students show *improvements* in their reflective skills through journal keeping. It was hypothesized that students' awareness of how they learn would improve as they progress through the course, engaging in continuous journal keeping. For instance, Dart et al. (1998) found that students' insights became more profound as their journals progressed, and the nature and quality of thinking and reflection, as well as their influence on practice, also developed.

Third, it was conjectured that students' writing skills would improve, that is, they were expected to make fewer grammatical and spelling errors and the readability and coherence of their journal responses was expected to improve. As suggested by Yinger (1985), journal keeping helps students to improve their writing by focusing on *processes* rather than on products, emphasising expressive and personal aspects, and serving as a record of thought and expression that is available for reading.

A final objective of the present study was attempting to code students' journal responses in an objective fashion by subjecting student journals to text analyses by means of an automated coding procedure, using software. By automating the coding process, the coding process is in principle performed in a consistent, objective and reliable manner, and can be performed in a fraction of the time required to do so manually. Furthermore, issues of inter-rater reliability are avoided, since the coding process is independent of the degree of interpretation by human coders. A last potential advantage of a coding scheme as developed in this study is that it can be easily applied to analyse large data sets of student journals. In order to maximize the validity of our findings, we did not rely on single journal entries of a selected, small group of students. Instead, the journals by all, more than three thousand first-year students of a polytechnic were involved in the study presented here with each of them writing approximately 150 journals in an academic year.

2. METHOD

2.1 Subjects

Participants were 3460 students in their first year of studies at a polytechnic in Singapore in the academic year 2007-2008. Of these students, 1765 (51%) were females and 1695 (49%) were males, and their mean age was 18.64 years ($SD = 1.46$).

2.2 Educational Context

Problem-based learning. The polytechnic at which the research was carried out organizes its curriculum according principles of problem-based learning. Students work collaboratively in teams of four to five, with learning centred on problems relevant to their domain of study. They work each day on one problem. The problem is initially discussed in the morning, followed by ample study. At the end of the day, information gathered is shared and elaborated upon. No didactic teaching takes place nor is there any form of direct instruction. One tutor supervises the student teams in a larger classroom. His or her role is to facilitate student learning (Alwis, 2007). There are two semesters in an academic year, with each semester lasting 16 weeks. All the courses offered are three-year curricula.

Assessment in the curriculum. Students' reflection journals form a part of the daily assessment approach. The reflection journal is a short essay created by the student that is "personal" and records his or her daily reflections of daily learning in respond to a reflection journal question provided by the tutor. Each student is required to submit his or her reflection journal by the end of the day. Tutor-asked journal questions mainly required students to be reflective about their learning and development. Some examples of journal questions include "What are some of the strengths that I demonstrated today?", "What insights did I gain today?", "What strategies have I used to help me in my learning", "What prior knowledge did I apply to help me understand the problem better?" and so on. Students respond to a different reflection journal question each day during a five-day workweek. The didactic purpose of writing

the reflection journal is in line with the literature reviewed above, to encourage and record self-reflection about the process of learning.

In addition, the daily assessment approach also involves students having to assess their own process of learning (i.e. self-assessment) and that of their peers (i.e. peer assessment), and a judgment by the tutor on how well students have performed during the day. Students also need to take four knowledge acquisition tests per module, which are taken at different points during the semester. The duration of each test is 30 minutes and it consists of answering at least three structured questions. The tests are conducted in a supervised environment, similar to an end-of-course examination. Students are tested on their ability to understand and apply what they have learnt.

2.3 Instrument

In seeking evidence of reflective activities through reflection journal writing, student journals were analyzed using the SPSS Text Analysis for Survey™ software (SPSS, 2006). The software uses advanced linguistic theory technologies that extract and classify key concepts from student journal responses. These technologies analyze content as a set of phrases and sentences whose grammatical structure provides a context for the meaning of a response. The software enables the coding and categorization of journal responses in a fraction of the time required to do the job manually. Another benefit is that the categorization of responses is done consistently and reliably; the responses are analyzed in an iterative manner. Unlike human coders, the software classifies the same response in the same categories every time.

The first step in content analysis is to extract key terms and ideas from the journal responses. The engine uses linguistic algorithms and resources to identify relevant concepts. This means that extraction does not treat a response as a set of unrelated words, but it identifies key words, compound words, and patterns in the text. The libraries supplied with the software contain pre-coded definitions were the linguistic resources used to extract terms from the journal responses.

The extracted terms were grouped into categories by the software. As used in content analysis, a category refers to a group of closely related concepts, opinions or attitudes. The software relies

upon three linguistic-based techniques that take into account the root meanings of the extracted terms and their relationship between sets of similar objects or opinions: *term derivation*, *term inclusion* and *semantic networks* (SPSS, 2006, p.101). Because these techniques are complementary to one another, all of them are used for categorizing the extracted terms.

The *term derivation* technique creates categories by taking a term and finding other terms that are related to it by analyzing whether any of the terms components are morphologically related. For instance, the term "opportunities for self-reflection" would be grouped with the term "self-reflection opportunities". The *term inclusion* technique uses algorithms to create categories by taking a term and finding other terms that include it. When determining inclusion, word order and the presence of such words as "in" or "of" are ignored. As illustration, given the term "skill", term inclusion will group terms such as "programming skills" and "a set of skills" in a skill category. The root term used to create the category (skill) can have words before it, after it, or both before and after ("programming skill set").

The *semantic networks* technique creates categories using a semantic/lexical network based on WordNet®, a linguistic project based in Princeton University (Miller, 2006). WordNet® is a reference system of "Nouns, verbs, adjectives and adverbs grouped into sets of cognitive synonyms, each representing one underlying lexical concept." This method begins by identifying extracted terms that are known synonyms and hyponyms (i.e., a word that is more specific than the category represented by a term, e.g., student, tutor and peer are hyponyms of the term "person").

In order to analyze the journal responses in a more meaningful fashion, a custom library was created. This library contained domain-specific words and terms (with synonyms) that arose from the modules taken by all first-year students. In this particular institution, all students were required to take two mathematics and computer applications modules in their first year of studies. These modules consisted of several tasks which asked students to create spreadsheets and basic computer programs to perform simple numerical functions. Using these modules as an example, domain-specific words would include "visual basics programming", "Microsoft

excel graphs”, “spreadsheets” etc. Subsequently, manual techniques (e.g. moving responses from one category to another and customization of the software’s libraries to generate more useful categories and to remove ambiguity) were used to provide finer control of the results. The categories that were automatically generated were also renamed to capture their essential meanings. The descriptions of the categories obtained are contained in Table 1.

In investigating whether students show improvements in their writing skills, we examined their spelling, grammar use, and overall readability and coherence of their journals. A Microsoft Word macro was created to detect the spelling errors present in student journals. The output generated gives a list showing the frequency count of each misspelled word in a given journal. The presence of grammatical errors in student journals was detected by means of the in-built grammar checker feature in Microsoft Word.

The two readability tests used in determining the reading level of student journals were the Flesch Reading Ease, and the Flesch–Kincaid Grade Level (Flesch, 1948; Kincaid, Fishburne, Rogers, & Chissom, 1975). In the Flesch Reading Ease test, higher scores indicate material that is easier to read. It rates text on a 100-point scale. For instance, scores of 90-100 are considered easily understandable by an 11-year old, and passages with results of 0-30 are best understood by college graduates. The Flesch–Kincaid Grade Level test rates text on a U.S. school grade level. For example, a score of 8.20 would indicate that the text is expected to be understandable by an average student in eighth grade.

Another Microsoft Word macro was created to determine the textual local coherence of student journals. *Local coherence* indicates the relatedness of subsequent sentences by *argument overlap*.

Table 1. Description of categories generated by means of text analysis software

Category	Sub-category	Description	Reference studies
Critical review of past learning experiences <i>Learners automatic awareness of the act of analyzing, evaluating and examining past behaviours and learning of self and others (peers), as well as the products of learning.</i>	Self	To look over or examine self-performance. This includes: <ul style="list-style-type: none"> • Being aware of one's capacity or weak points to learning effectively, i.e., strengths and weaknesses, • Setting or tracking learning goals, • Manner in which students consistently respond to and use stimuli in the context of learning i.e.: learning styles such as visual (learn best through visual displays), auditory (learn best through listening) and tactile (learn best through hands-on approach) 	(Lew & Schmidt, 2006; Lew & Schmidt, 2007a, 2007b; Moon, 1999a, 1999b)
	Peers	To look over or examine peers' performance. This includes: <ul style="list-style-type: none"> • Team work, and team dynamics, i.e. cooperativeness and level of contributions, and, • Helping peers with their learning, or seeking help from peers. 	
	Products	To look over or study the products of learning, which emerged as a result of relating knowledge structures from text. This includes: <ul style="list-style-type: none"> • Domain-specific skills, e.g.: graph-plotting using Microsoft Excel, Visual Basics programming, and Microsoft PowerPoint etc. • Presentation slides, self-created computer programs, self-creating Excel accounting spreadsheets, classroom performance grades etc. 	
Learning strategies <i>Ways in which students use to plan their learning, as well as various methodologies used.</i>	Rehearsal	Oral repetition, copying, making selective verbatim responses and underlining the important parts of the material	(McCombs & Whistler, 1989;
	Organization	Categorizing information, creating knowledge networks and hierarchies (e.g. mind maps)	Weinstein & Mayer, 1986)
	Elaboration	Creating analogies or mental images, <u>generative note taking and self-questioning.</u>	
	Others	Focusing attention and motivation, managing performance anxiety and time management	
Summaries of what was learnt	-	Student restate in their own words, what they had learnt. They generate narratives of their own experiences, learning and development. For instance, relating new information to prior or existing knowledge; applicability of knowledge gained to other situations.	(Selfe et al., 1986; Woodward, 1998)

Local coherence is found when the second sentence contains an idea previously mentioned in the first sentence (Britton & Gilgoz, 1991). All the macros used in the analyses of student journals were created using Visual Basic for Applications (VBA), an event-driven programming language which is built into Microsoft Word. More details regarding the macros can be found in Appendix E.

2.4 Procedure

Data used in the analyses were student reflection journals for the entire week, collected once at the beginning of (i.e. Week 3 of the first semester), and again, at the end (i.e. Week 14 of the second semester) of the academic year 2007-2008. Identical categories were generated for both sets of data. The number of instances which each category appeared in each journal response was recorded and used for subsequent statistical analyses.

All the journals were subjected to spelling, grammar, readability, and coherence tests using VBA macros created in Microsoft Word.

2.5 Analysis

Descriptive statistics (means and standard deviations) of the frequency counts for the categories were computed. Paired-samples *t* tests were also performed to examine if the differences in the mean frequency count of the categories generated based on student reflection journals written at the beginning and the end of the academic year were statistically significant.

The total counts of the number of spelling errors and that of the grammatical mistakes for the journals written by each student were computed. Similarly, the readability and coherence test scores were also recorded. Paired-sample *t*-tests were then performed on these scores and reported.

To further examine the magnitude of the difference in the mean categorical values for that of week 3 as compared to that in week 14, effect size (Cohen's *d*) values were computed. A rule of thumb for describing the magnitude of effect sizes can be attributed to Jacob Cohen (1969). According to Cohen, he suggested that an effect size

of 0.20 should be regarded as “small”, 0.50 should be regarded as “medium” effect size, and an effect size of 0.80 should be regarded as “large”. Furthermore, a positive effect size represents improvement whilst a negative effect size indicates deterioration.

3. RESULTS

Descriptive statistics for the number of instances with which each textual category generated by text appeared in student reflection journals are given in Table 2. The findings suggest that students reflected on three general categories related to their learning in their journal responses: critical review of past learning experiences, learning strategies and summaries of the contents of what was learnt. Students appeared to focus most on reflecting on their learning behaviours and performance. Furthermore, students seemed to focus least on reflecting the content of what they have learnt, as indicated by the low frequency counts of the ‘summaries of what was learnt’ category. The means between the categorical frequency counts in student journal responses obtained in weeks 3 and 14 were tested for significant differences using paired-samples *t* tests and the results contained in Table 2. The computed effect sizes are also contained in this table.

The outcomes of the analyses demonstrate that the differences between the two data sets were statistically significant. Absolute Cohen’s *d* values ranging from 0.16 to 0.80 were obtained, suggesting small to large effect sizes. Furthermore, a mixture of both positive and negative effect sizes were obtained based on the magnitude in the difference in categorical means, indicating that students appeared to reflect and write more about certain aspects of their learning in their journals, though less so in other areas as they progressed through the academic year.

Table 2. Descriptive statistics, outcomes of paired-sample *t* tests and effect sizes of frequency counts for categories present in student journal responses

	Week 3		Week 14		Paired-samples <i>t</i> tests	Effect size
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>t</i> value	Cohen's <i>d</i>
Critical review						
Self	4.40	2.93	6.42	2.04	4.43**	0.80
Peers	6.50	2.15	5.43	2.48	5.62**	-0.46
Products	5.79	3.48	8.73	4.44	5.30**	0.74
Learning strategies						
Rehearsal	4.02	1.33	5.74	2.91	4.38**	0.76
Organization	2.87	1.86	3.89	1.35	8.41**	0.63
Elaboration	1.34	1.91	2.37	1.88	5.58**	0.54
Others	3.60	1.49	3.36	1.55	6.93**	-0.16
Summaries of what was learnt	4.36	2.81	2.87	1.15	4.34**	-0.69

Note. *SD* = standard deviation.

Degrees of freedom = 3414

***p* < .01, 2-tailed

The descriptive statistics for the spelling, grammar and readability tests are given in Table 3.

The results demonstrated that as students progressed through the academic year, the reading ease of their journal responses increased, whilst the Flesch–Kincaid Grade Level decreased. On the other hand, there were more spelling and grammatical mistakes in the reflection journals written by students at the end of the academic year as compared to the start of the year. The outcomes of paired-sample *t* tests demonstrate that the differences between the mean values were statistically significant. Absolute Cohen's *d* values ranged from 0.10 to 0.35, suggesting small effect sizes. Finally, the difference in the mean coherence values was not statistically significant, indicating that the writing style of the students did not become more coherent over the year.

Table 3. Descriptive statistic, outcomes of paired-sample *t* tests and effect sizes for spelling, grammar and readability tests

	Week 3		Week 14		Paired-sample <i>t</i> test <i>t</i> value	Effect size Cohen's <i>d</i>
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>		
Spelling mistakes	4.21	8.34	6.64	5.58	12.24**	0.34
Grammatical mistakes	9.73	7.25	12.17	6.54	5.71**	0.35
Readability						
Flesch Reading Ease	61.65	15.62	65.43	12.89	2.06**	0.26
Flesch–Kincaid Grade Level	11.25	6.77	10.17	8.83	5.81**	0.14
Coherence	.43	.17	.41	.24	0.49	0.10

Note. *SD* = standard deviation.

Data used were student reflection journals from week 3 of the first semester and week 14 of the second semester in the 2007-2008 academic year.

Degrees of freedom = 3414

***p* < .01, 2-tailed

4. DISCUSSION

The present study was conducted to examine student journals for evidence of reflective activities, and whether students show improvements in their reflective skills through continuous engagement in journal writing as they progressed through the academic year. To that end, we attempted to code the reflection journals written by students in an objective fashion, by means of an automated content analysis approach using software. We were interested to find out whether students' journals, in agreement with what the literature suggests, focus on critical reviews of their own and/or that of their peers' processes of learning and behaviours, and understanding of learning strategies used to enhance their metacognition. Furthermore, we were also interested to find out to what extent students use reflection journals to summarize the content of what they have learnt. This is considered by some as ineffective in enhancing students' reflective abilities, as it is thought to hinder the objective of developing critical thinking and metacognitive skills (Langer, 2002; Voss, 1988).

The outcomes of the text analyses suggest that there is evidence of reflection in students' reflection journals; they appeared to reflect on three general categories related to their learning: critical review, learning strategies, and summaries of what was learnt. Descriptive statistics of the mean categorical counts suggest that students focused most on reflecting on their learning behaviours and performance. Furthermore, students seemed to focus least on reflecting on what they have learnt at particular points in time. In addition the number of reflective activities increased while the year progressed as indicated by the significant differences between the mean frequency counts for the three general learning categories found in student journal responses. This suggests that some learning took place during the course of the academic year, such as in the area of enabling students to become better aware of their learning. Students appeared to show more engagements in critical reviews about the processes of their own learning and behaviours, and demonstrated a better understanding of their ability to transform their own learning strategies. This concurs with what the literature reports about the role of reflection journals in enhancing students' awareness of their cognitive processes and their control of these processes (McCrinkle & Christensen, 1995; Vockell, 2004). By contrast, students showed fewer tendencies to summarize the content of what they had learnt, as indicated by the significant decrease in the mean frequency count for the category on "Summaries of what was learnt" as the academic year progressed. This suggests that some learning took place, since students restating in their own words what was learnt are considered by some as ineffective in enhancing students' reflective abilities, as it is thought to hinder the objective of developing critical thinking and metacognitive skills (Langer, 2002; Voss, 1988).

What do these findings imply? First, there is the possibility that our results were somewhat coincidental despite finding significant differences. The reader may remember that students write reflection journals in response to a question of their tutor. These questions differ per day. Some examples of tutor-asked journal questions include "What are some of the strengths that I demonstrated today?", "What insights did I gain today?", "What strategies have I

used to help me in my learning”, “What prior knowledge did I apply to help me understand the problem better?” and so on. They also differ between tutors. Therefore, the differences between responses in week 3 and 14 may be caused by differences in the particular questions asked. To test whether the difference in reflection on learning as a function of time was influenced by the specific tutor-asked questions, we subjected all questions asked in both week 3 and 14 to text analyses using the same content analysis approach of student journal responses. In total, more than 1000 journal questions were asked by approximately 250 tutors involved in taking first-year classes. Identical categories (e.g. learning strengths and weaknesses, learning goals, collaborative learning etc.) were generated for both data sets. Comparisons between the means of the frequency counts for the categories by means of paired sample t tests revealed that none of their differences were statistically significant. Therefore, the increase in the amount of reflection in students’ journals cannot be explained away by differences in questions asked.

A second possibility is that the findings are time-dependent, *viz.*, the results obtained would have been different if student journal responses from other weeks of the academic year would have been used in the content analyses. To test this hypothesis, we examined *post-hoc* the journal responses of students written in two other weeks, i.e. week 4 of the first semester and week 15 of the second semester. Identical categories to those contained in Table 1 were generated. Similar to the results obtained from the data sets from weeks 3 and 14, test of differences between the mean categorical frequency counts by means of paired-samples t tests revealed no significant differences (for example, Critical review (self) = $t(3459) = 4.47, p < .01$; Learning strategies (organization) = $t(3459) = -8.37, p < .01$; Summaries of what was learnt = $t(3459) = 4.65, p < .01$, with degrees of freedom in parentheses). This suggests the measurement stability of our findings, since the results from content analyses using data from other weeks of the academic year were similar to those obtained from the data sets from weeks 3 and 14.

A third factor potentially affecting our findings is that the presence of spelling and grammatical errors may have influenced the outcomes of the resulting categories to some extent. An increase in

the number of errors over time may have increased that number of responses per category by the text analysis program. If, for instance, students wrote shorter sentences to a larger extent in week 14 (as suggested by the Flesch–Kincaid grade level readability score), and these shorter sentences were grammatically more often incorrect because they were shorthand rather than full sentences, the number of responses may have increased. This arises because of a limitation of the text analysis software, which could not detect the presence of wrongly spelt words. For instance, misspelt terms such as “teamwk” (instead of “teamwork”) and “grp” (instead of “group”), were extracted and counted towards the categorical count of Critical review (peers). To test this hypothesis, a random sample of 500 first-year students was first selected. The spelling errors in the reflection journals (over 1000 responses from week 3) of these students were first corrected for spelling errors before they were subjected to text analyses. The analyses generated identical categories to those contained in Table 1. Test of differences between the mean categorical frequency counts by means of paired-samples t tests revealed no significant differences between the data set from this random sample and that from week 3 (for example, Critical review (self) = $t(499) = -1.29, p < .01$; Learning strategies (organization) = $t(499) = 0.31, p < .01$; Summaries of what was learnt = $t(499) = 0.76, p < .01$, with degrees of freedom in parentheses). Therefore, writing errors do not seem to affect the outcomes of the text analyses.

In summary, our findings indicate that there is evidence of reflection in student journal responses, in accordance with what several authors emphasized that reflection journals, by focusing on the *processes* (e.g. self-reflection and learning strategies) rather than the *products* (e.g. summaries of knowledge acquired) of learning, may enable students to improve on their reflective capacity and skills (Dart et al., 1998; Haigh, 2001; Voss, 1988). The findings further suggest that students show improvements in their reflective skills through journal keeping. Students’ awareness of how they learnt improved as they progress through the academic year as they engaged in continuous reflection journal writing. Similar results were also mirrored in the study by Dart et al. (1998). Although the findings from the present study seem to concur with the literature on

reflection journal writing, the results from existing studies were more subjective, since they involved manual coding of student journal responses done in a non-standardized manner. In addition, existing studies did not include comparison of findings over time, casting some doubts over the reliability and validity of their results.

Contrary to most studies in journal writing with limitations such as small sample size, non-continuous engagement in the task of writing journals or infrequent feedback given by teachers, the present study has sought ways in arriving at more reliable and valid measurements. We did not rely on single reflection journals of students and had adopted an automated coding procedure, where the categories were derived based on findings in the literature of the positive effects of journal writing on student learning. As such, the problem of inter-coder reliability was absent. Furthermore, in this context, students got continuous feedback on their behaviours and learning from their tutors. Such the provision of regular and timely feedback may have created optimal conditions for enhancing students' awareness of their how they learnt, given that they receive explicit cues and suggestions from their tutors on how they can further improve on their learning.

Changes in writing skills while producing reflection journals throughout the year were also examined, in particular spelling, grammar, readability, and local coherence. Outcomes of paired-sampled *t* tests suggest that students wrote simpler reflection journals with higher readability as they progressed through the academic year. In addition, spelling and grammar became poorer. This suggests that despite constant engagement in the task of reflection journal writing, writing ability does not improve. Students' written expression of ideas in a more coherent manner did improve even after engaging in journal writing for a year or so.

4.1 Limitations

Some limitations should however be noted. The text analysis software is not a panacea, and although using software to perform content analysis removes inter-coder reliability as a concern, it is not without its shortcomings. In human coding, the coders read the

responses and can capture all the nuances of a statement even if they face difficulties applying the coding categories. The software can apply the coding categories, but they need to be defined so that the nuances are captured. An implication arising from this is that the editing done by the researchers of the synonyms and excluded words in the various libraries must accurately capture the ideas of the respondents in the text. Another limitation of the software is that it will not capture all the information in the journal responses, although categories can be created easily without any intervention on the part of the researchers. In examining readability and coherence of student journal responses, as the readability tests used do not factor in the meaning of words in a given text, they are thus not definitive measures of reading ease. To add on, the coherence test used could only give a crude approximation of whether the meanings and sequences of ideas relate to one another in a given text, and is also not a definite measure of textual coherence.

4.2 Future Research

Two other issues present themselves for future research based on the findings from the present studies. First, given that there is evidence of learning in student reflection journals and if students do benefit from the activity of journal keeping, their academic performance is expected to improve. Thus, examining the effects of journal writing on academic achievement should constitute the next stage of research. Second, further research should investigate if student writing skills can be improved through formal training in journal writing. Through a more structured and closely guided process, not only will students' ability to write better improves, they also become better metacognitive and self-reflective learners, a skill so valued in professional practice.

Chapter 6- Summary of research findings, implications and future directions

The preceding chapters reported studies on student self-assessment in higher education, the theme of this thesis. In Chapter 1, I presented a literature review on existing research which examined the effects of self-assessment on student learning, and empirical studies which examined student self-assessment accuracy. I also highlighted the relevance of studying it, since many researchers and educational practitioners have acknowledged the positive role that self-assessment may play in student learning, and in the development of professional competence. In that chapter, I also discussed reflection journal writing as an activity that may provide students with opportunities for self-assessment. In addition, Chapter 1 introduced the research questions studied in this thesis. These questions were: (1) How do students and teachers differ in their views about the purposes and utilities of self-assessment tools?; (2) What are students' beliefs about the utility of self-assessment?; (3) How accurate are students' self assessments as compared to peer and tutor assessments?; and (4) Does reflection journal writing improve student learning? The next chapters explored these topics. Chapter 2 reported findings of a focus-group study with students and tutors aimed at examining the first question. The validation study of a questionnaire to address the second question is presented in Chapter 3. Chapters 4 and 5 focused on answering the third and fourth research question respectively. In this chapter, I initially turn back to the research questions to summarize my findings. Subsequently, I discuss implications of these findings to higher education, and finally, suggest directions for further research. I will however start with a brief description of the learning environment in which the studies took place and the actual assessment procedures used in that environment.

6.1 The Republic Polytechnic problem-based curriculum

The polytechnic at which the research was carried out organizes its curriculum according to principles of problem-based learning. Here,

students work collaboratively in teams of four to five, with learning centred on problems relevant to their domains of study. Students work each day on one problem during a five-day work week. The problem is initially discussed in the morning, followed by ample study. At the end of the day, information gathered is shared and elaborated upon. No didactic teaching takes place or is there any form of direct instruction. One tutor supervises the student teams in a larger classroom.

Assessment at the polytechnic involves students being graded daily, and they having to take knowledge acquisition tests. The daily assessment approach consists of four, independent elements: (1) a self-assessment, (2) a peer assessment (3) a reflection journal, and (4) a judgment by the tutor on how well students have performed during the day. The self-assessment rating scale consists of 8 items inquiring about the quality of students' performance within their team, such as the level of cooperativeness and contribution of ideas. A Cronbach's alpha value of .90 gives evidence for the high internal consistency reliability of the self-assessment instrument. The peer assessment rating scale consists of 4 items inquiring about the cooperativeness and quality of contributions of peers within the team. The peer assessment instrument has high internal consistency reliability, given its Cronbach's alpha of .93. In examining the inter-rater agreement by correlating the scores awarded to students by different peers, we computed the intraclass correlations based on students' peer assessment scores. Intraclass correlations of .97 and .95 for the first and the second semester respectively were obtained. The values of Cronbach's alpha were computed based on student responses on the items of the self- and peer assessment instruments in semester one of the 2007-2008 academic year. Students are asked to respond to these items on a Likert five-point scale ranging from "strongly agree", "disagree", "neutral" and "agree" to "strongly agree". The items for the self- and peer assessment are contained in Appendix A. On a particular day, each student assesses and is in turn assessed by his peers within the team.

Students' reflection journals form a part of the daily assessment approach. The reflection journal is a short essay created by the student that is "personal" and records his or her daily reflections of

daily learning in respond to a reflection journal question provided by the tutor. Each student is required to submit his or her reflection journal by the end of the day. Tutor-asked journal questions mainly required students to be reflective about their learning and development. Some examples of journal questions include "What are some of the strengths that I demonstrated today?", "What insights did I gain today?", "What strategies have I used to help me in my learning", "What prior knowledge did I apply to help me understand the problem better?" and so on. Students respond to a different reflection journal question each day during a five-day workweek. The didactic purpose of writing the reflection journal is in line with the literature reviewed above, to encourage and record self-reflection about the process of learning.

The tutor judgment consists mainly of tutors' observations of students' processes of daily learning. The observations by the tutors include students' self-directedness, level of participation inclusive of teamwork; students' ability to reason, justify and defend opinions and ideas formulated in respond to problems, as well as their problem solving skills. Tutors will then award grades ranging from "A" to "F", which are derived based on what they observe and the impression they have on each student during the duration of time they had with him/her. Tutors also take into consideration students' individual reflection journals (short essays which document students' reflections on daily learning) and their self and peer assessments when awarding grades. Furthermore, tutors will provide feedback to students on their learning outcomes and processes of daily learning. The generalizability of judgments made by different tutors is high, with an average generalizability coefficient of .84 (Chua & Schmidt, 2007).

Students also need to take four knowledge acquisition tests per module, which are taken at different points during the semester. The duration of each test is 30 minutes and it consists of answering at least three structured questions. The tests are conducted in a supervised environment, similar to an end-of-course examination. Students are tested on their ability to understand and apply what they have learnt.

6.2 Summary of findings

1. *How do students and tutors differ in their views about the purposes and utilities of self-assessment tools?*

Chapter 1 describes what is presently known about alternative assessment in higher education. Alternative forms of assessment, for instance, self-assessment, peer assessment and reflection journals were introduced because of their ability to test for and measure higher order competencies, such as metacognitive and interpersonal skills (Falchikov, 2005; Segers & Dochy, 2001). They seek not only to encourage reflection, the ability to evaluate the performance of oneself and one's peers, but also serve to actively engage students in their learning process. Nonetheless, some researchers have reported problems associated with alternative assessment, such as concerns of students over the fairness and authenticity of such forms of assessment (e.g. Struyven, Dochy, & Janssens, 2002), students having underdeveloped conceptions of what assessment was, and teachers were using the assessment forms to rank and grade students, instead of using them as feedback tools to help students further improve on their learning (MacLellan, 2001). Kerka (1996) also reported that students use their reflection journals as tools to criticize fellow students.

The use of alternative assessment tools assumes that both teachers and students understand what the *raisons d'être* of these instruments are. We were interested to examine the extent in which a curriculum-wide introduction of assessment aimed at self-reflection causes the problems described above to arise. To that end, we conducted a focus-group study which compared teachers' and students' views about the purposes and utility of assessment tools in a post-secondary institution which organizes its curriculum according to principles of problem-based learning. Tutors ($n = 7$) and students ($n = 15$) were asked to share their views about the self- and peer assessment activities, reflection journals, and classroom performance grades (i.e. tutor judgment). Chapter 2 reports the findings of this study.

The findings of the focus-group study suggest that both teachers and students understood the purposes of self-assessment but perceptions of its actual use differed. Teachers generally believed that a multifaceted approach provides a rich understanding of how well students are learning, and self-reflection may help students to become better learners. Teachers were however sceptical about using students' self- and peer assessment scores to corroborate against their judgments of students' performance when grading. By contrast, students could not see the various assessment instruments as valuable in their own right (as was the purpose). They preferred to reflect on content rather than on the process of learning in their reflection journals. Many of them believed that their self assessments and journal responses could be used to influence the teachers' impressions of their performance, and were used by their teachers (to some extent) to arrive at their final classroom performance grades.

Two reasons were mentioned to explain the results. First, students assessed themselves better than they actually performed in their self assessments. They assumed that by doing so, they could impress their teachers, which may warrant better classroom performance grades. Students also appeared to let interpersonal relationships with friends interfere in their peer judgments; even if their peers contributed minimally towards teamwork, they could not care less to assess their peers accordingly in their peer assessments. Second, students believed that writing qualitatively good journal responses improved their classroom performance grades. The perceived influence of the reflection journal on students' grades may be in its ability to reflect on personal learning achievements, and to account for individual effort in a collaborative learning environment like problem-based learning.

2. *What are students' beliefs about the utility of self-assessment tools?*

Self-assessment has been conceived as learners' ability to make appraisals of aspects of their learning. It is assumed that through assessing their own performance, students engage in metacognition

which enables them to reflect on their own accomplishments, to monitor their progress while learning, and to internalize standards of performance so that they can regulate their learning more effectively (Dochy, Segers, & Sluijsmans, 1999; Segers & Dochy, 2001). Self-assessment is not only expected to encourage self-reflection or the appraisal of one's abilities, it is also supposed to actively engage students in their own learning process. These features of students' learning are crucial in assisting them to become self-regulated, life-long learners who develop control over their own learning (Mok et al., 2006; Paris & Paris, 2001).

Existing studies in the literature which compared students and teachers' views about self-assessment were largely based on *ad hoc* questioning of them about its utilities. The absence of a validated instrument in the literature to measure students' beliefs about the utility of self-assessment tools was the motivation behind the three studies reported in Chapter 3. By building upon relevant literature and insights collected from students and teachers through focus-group interviews, I conceived a theoretical model of students' beliefs about the utility of self-assessment tools in higher education. It is suggested that students' beliefs can be classified into seven categories. Students' beliefs about reflection journal writing were represented by four belief categories: (I) *The reflection journal enabled me to think and write reflectively.*; (II) *Frequent journal writing improves my learning.*; (III) *I can look good in front of my tutor when I write a qualitatively good reflection journal.*; and (IV) *The reflection journal enables me to provide feedback to my tutor about my peers' performance.* For the self-assessment activity consisting of Likert-scale items, three belief categories suggested were: (V) *The self-assessment enables me to make an appraisal of my learning.*; (VI) *Assessing my own performance is more of a habitual action than to improve on my learning.*; and (VII) *The self-assessment enables me to manage my tutor's impressions of how I performed.*

In the first study, a 31-item questionnaire containing seven factors based on belief categories derived in the literature as well as findings from Chapter 2 about the utility of self-assessment on student learning was developed. It was administered to a representative group of 327 second-year students, and results were

analysed using a structural equation modelling approach. This approach provided a set of relevant statistics - Chi-square, degrees of freedom, the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA) - that were used to verify whether the hypothesized model explained the empirical data. Coefficient H values for each factor were also computed to determine construct reliability. The revised questionnaire containing 25 items of self-reflection beliefs was tested using confirmatory factor analysis. It was administered to a second, independent sample (273 second-year students) from the same student population to cross-validate the proposed model. Furthermore, in investigating if the questionnaire model in the first two studies replicates across independent samples of the same population, test of measurement invariance was conducted across the two samples simultaneously. The second study reports these findings. The third study examined the measurement stability of students' beliefs over time; *viz.*, the validated 25-item questionnaire from the first and second studies was assessed for its test-retest reliability. Correlational analyses were also conducted to examine the inter-relationships between the belief categories.

In summary, the findings from these studies suggest that students were able to identify the seven latent constructs underlying the questionnaire, as indicated by the fit of the hypothesized model. The test for measurement invariance showed that factor loadings were equivalent across different student groups and the questionnaire's underlying structure gave evidence of cross-validation. Evidence for sufficient test-retest reliability was also found suggesting stability of beliefs over time. These findings taken together demonstrate that the questionnaire developed appears to be an adequate instrument for measuring students' beliefs about the utility of self-assessment on their learning. Factor correlations demonstrate that students believe that self-assessment can have multiple purposes, including self-improvement and impression management of teachers that are not necessarily in accordance with each other.

3. *How accurate are students' self assessments as compared to peer and tutor assessments?*

The upsurge of interest in student self-assessment among researchers and educators arises from the recognition of the possible positive role that self-assessment may play both in learning and in the development of professional competence (Boud, 1989; Sluijsmans, Moerkerke, & Dochy, 1998). Besides studies which have examined students' beliefs about self-assessment, and its effects on student learning, there exists a body of literature reporting empirical studies that compare student-provided marks with those of teachers. In light of this type of self-assessment, research usually looks into the validity of the grades, by comparing the accuracy of the grade given by the learner with that given by the teacher or their peers (Boud & Falchikov, 1989; Falchikov & Boud, 1989). Nonetheless, most of the self-assessment studies in higher education focus on student assessing their capacity to acquire content knowledge, and of the accuracy of their self-predictions of performance when compared with actual achievement. However, less is known of students' ability to make judgments about their own learning *process*, *viz.*, the act of self-monitoring their learning development, identifying strengths and weaknesses, and adapting learning in light of experience and feedback from teachers and peers. Self-assessment takes on this latter interpretation for the two studies reported in Chapter 4. This chapter seeks to evaluate the accuracy of students' self-assessment ability, to examine whether this ability improves over time, and to investigate whether self-assessment is more accurate if students believe that it contributes to improving learning.

In the first study, the accuracy of the self assessments of 3588 first-year students enrolled in a post-secondary institution was studied throughout a semester during which each student made approximately 80 self assessments about his or her own learning process. These self assessments were then compared with multiple judgments by peers and tutors. The overall correlations between the scores of self, peer and tutor assessments (r ranging from $-.03$ to $.31$) suggest weak to moderate accuracy of student self-assessment ability. The findings also reveal an ability effect; students judged as

more academically competent were able to self-assess with higher accuracy than their less competent peers. The correlations between students' self- and peer assessment scores are higher for high-achieving students as compared to low-achieving students ($r = .41$ and $.23$ respectively). Similarly, the correlation between students' self-assessment scores and tutor grades for high-achieving students is higher for high achievers as compared to low achievers ($r = .29$ and $.01$ respectively). The difference between these r -values in both cases were statistically significant ($p < .01$). Comparing the accuracy of student self-assessment averaged over four consecutive periods demonstrated a gradual, decreasing trend in the linear relation between the self- and peer assessment scores is observed, with moderate r -values ranging from $.37$ to $.29$. A similar change pattern is noted for students' self-assessment scores and their tutor judgments, with low r -values ranging from $.27$ to $.15$. The correlations between students' self-assessment scores and tutor grades are comparatively lower compared to students' self-judgments and the judgments by their peers. Results of test of differences between correlations drawn from the same sample revealed that the differences in the correlations between self- and peer assessment scores, and that of self- and tutor scores computed for different time intervals were not statistically significant. Taken together, our findings indicate that students on average do possess accurate self-assessment skills only to a limited extent. In addition, our studies provide evidence that self-assessment is not learned through extended experience and regular feedback from teachers and peers.

In the second study, the self-assessment section of the validated questionnaire from Chapter 3 was administered to 936 first-year students. Based on their responses, sub-groups of students were identified: those who either believed in the usefulness of self-assessment or did not. Outcomes of correlational analyses demonstrate that moderate relations exist between the self- and peer assessment scores for those students who believe that the self-assessment is useful as an impression management tool ($r = .43$) as compared to those who believe otherwise ($r = .32$). The results also show that low relations exist between self-assessment scores and tutor grades for those students who hold strong beliefs that self-

assessment is useful as an impression management tool and the scores of those students who hold weak beliefs ($r = .19$ and $.21$ respectively). A moderate inter-relationship exists between self-assessment scores and tutor grades for students who believed in the usefulness of the self-assessment as a learning but not as an impression management tool ($r = .42$). By contrast, no significant relation between these scores for students who believed in the usefulness of the self-assessment for impression management and not for learning was reported. Testing for differences among the correlations between self-assessment scores and tutors' grades for the various student groups indicated that these were not statistically significant, suggesting that the accuracy of self-assessment is no different for students who hold strong beliefs and for those who hold weak beliefs about the utility of self-assessment on their learning. So, whatever students believe about the effects of self-assessment on their learning, no outcomes can be observed on their self-assessment accuracy.

4. *Does reflection journal writing improve student learning?*

Self-reflection has been highlighted as an important requirement for professional competence (Boenink et al., 2004). The ability to reflect on one's own knowledge and experience has been pointed as a means of dealing with the complexities, challenges and uncertainties inherent in professional life (Moon, 1999a; Thorpe, 2004). Research which looks into the use of reflection journals as learning and assessment tools to encourage self-reflection, cognitive and metacognitive processes of student learning has received much attention alongside the introduction of reflective practice in many institutions of higher education. Reflection journals provide many opportunities for students to engage in self-assessment, since the learner is engaged in processes which explore experience as a means of deepening understanding (Boud, Keogh, & Walker, 1985; Moon, 1999b). The literature offers evidence that students, regardless of their domains of study, show improvements in their learning, viz., students becoming better in self-assessment through journal writing (Moon, 1999a; Selfe, Petersen, & Nahrgang, 1986). Despite such

theoretical justifications, empirical evidence in support of this is still scarce. Furthermore, existing empirical studies which examine the level of reflecting thinking in students' journal responses by means of coding schemes have several limitations (Boud et al., 1985; Kember et al., 1999; Wong et al., 1995). First, these studies generally involved only a limited number of participants. There are therefore issues with the validity and generalizability of the findings. Second, manual coding of student journal responses is time-consuming and tedious, and has cast doubts over inter-coder reliability. Third, the coding procedures described in these studies lacked details on how they were carried out, or were too complicated for use in analyzing large number of student journal responses.

The study described in Chapter 5 was conducted to see whether the shortcomings associated with the existing empirical studies mentioned above can be overcome. To that end, I attempted to code the journal responses by over 3000 first-year students using an automated procedure by means of software. The data was collected twice over a period of an academic year: once at the beginning of the year and again towards the end of the year. During this period, each student wrote approximately 150 reflection journals. The outcomes of the text analyses demonstrate that there is evidence of student learning in their journal responses. They appeared to reflect on three general categories related to their learning: critical review, learning strategies, and summaries of learning. *Critical review* refers to the act of analyzing, evaluating and examining past, present and/or future behaviours and learning of self and others (peers), as well as the products of their learning. By *learning strategies*, we were referring to the ways in which students use to plan their learning, as well as various methodologies used to enhance the manner they learn. The reflective category on *summaries of learning* seeks to measure the number of instances in which students restated in their own words, what they had learnt. This, however, is considered ineffective in enhancing students' reflective abilities as it hinders the objective of developing critical thinking and metacognitive skills (Langer, 2002; Voss, 1988).

Descriptive statistics of the mean categorical counts suggest that students focused most on reflecting on their learning behaviours and

performance. Furthermore, students seemed to focus least on reflecting what they have learnt. Effect size (Cohen's d) values were computed in examining the magnitude of the difference in the mean categorical values for that at the beginning of the academic year as compared to those at the end of the academic year. Absolute Cohen's d values ranging from 0.16 to 0.80 were obtained, suggesting small to large effect sizes. Furthermore, a mixture of both positive and negative effect sizes were obtained based on the magnitude in the differences in categorical means, indicating that students appeared to reflect and write more about certain aspects of their learning in their journals, though less so in other areas at different periods of the academic year.

We also examined whether student writing skills improve with frequent engagement in journal writing through performing a series of tests: spelling, grammar, readability and local coherence. Outcomes of paired-sampled t tests suggest that students wrote simpler reflection journals with higher readability and poorer coherence as they progressed through the academic year. This suggests that students were unable to improve on the written self-expression of their ideas in a more coherent manner even after engaging in journal writing for a year or so. The reading ease of their journals also decreased. Taken together, the findings of the two studies presented here seem to suggest that journal writing improves student learning; students' awareness of various aspects of their learning increased. In addition, our studies provide evidence that students' writing skills are not learnt through extended experience.

6.3 Implications

The following paragraphs present the implications emerging from these research findings to effects of classroom applications of self-assessment on student learning, present within existing literature over the last three decades, and to higher education.

First, the qualitative findings from the study emphasize the importance of striving to obtain a match between the didactic functions of different assessment measures, and how students perceive and utilize them. Several researchers (e.g. Langer, 2002;

Segers & Dochy, 2001) have compared teachers' and students' views about various assessment forms and reported mismatch between their perceptions. My own study, discussed in Chapter 2, reported similar results. In addition, the findings highlighted that the students' perceptions and scepticism of self-, peer and tutor assessment may negatively influence the objective of developing higher order competencies such as cognitive and interpersonal skills. Simply getting students to reflect on their own and that of their peers' learning and behaviours may not improve their learning. The factor correlations reported in Chapter 3 showed that whilst students saw the value of the self-assessment tools on their learning, they did not exploit their benefits to help them to improve, but respond to these tasks in a manner which they thought will assist them in achieving better classroom performance grades.

The findings reported in Chapters 2 and 3 when taken together, points to the need for distinction between assessment tools used for *formative* and *summative* purposes, and explicit instructions to students on the didactic purposes and functions of these assessments before engaging in them. These implications emerging from the research may provide guidance for the introduction of alternative assessments (e.g. self- and peer assessment activities) in classrooms. This could contribute to appropriately identifying ways of exploiting the benefits of self-assessment on students' processes of learning and behaviours, and to enhance these effects on their academic performance.

Second, the empirical evidence arose from the study reinforces findings that students' ability to self-assess accurately is not learned through extended experience and regular feedback. Self-assessment studies which examined the student self-assessment accuracy by comparing their self-judgments with those by teachers and peers demonstrates that student self-assessment accuracy does not improve with increased seniority in the course of study (Fitzgerald, White, & Gruppen, 2003). Furthermore, the provision of regular feedback from teachers and peers on students' performance does not improve students' ability to self assesses (Eva et al., 2004). Empirical evidence also supports findings linking self-assessment skill and intellectual capacity. It was repeatedly reported in the literature

that students judged as being more competent academically are able to self-assess with greater accuracy, given that they are better at self-monitoring, judging their own performance and processes of learning, and at identifying their own learning strengths and weaknesses (Boud & Falchikov, 1989; Orsmond, Merry, & Reiling, 1997b). The study discussed in Chapter 4, demonstrated that high-achieving students have better developed self-assessment skill, making them more competent self-assessors as compared to low-achieving students.

An implication which arises from the study reported in Chapter 4 is that students' extended engagement with the self-assessment activity, and them receiving regular feedback on their performance from teachers and peers are not enough to affect change. The provision of a structured and closely guided process which trains students to self-assess may improve their ability in this area. Furthermore, empirical evidence which may shed light on characteristics and factors which could account for individual differences in student self-assessment skill is lacking. These two aspects will be further explored in the discussion of future directions for research, after consideration of other implications of my studies.

Third, findings reinforce the importance of development of self-reflection in students.

6.4 Further Research

Self-assessment has called the attention of researchers concerned with its potential benefits to improving student learning. This dissertation may have contributed to this endeavour by bringing light to the utility of self-assessment as seen through the eyes of teachers and students, the ability of students to self-assess accurately as compared to judgments by teachers and peers, and effects of journal writing on improving student learning. Those engaged on classroom applications of self-assessment may also benefit from questions emerging from this thesis. Unsolved issues came out from the studies and indicate directions for future research.

A first question to be further explored refers to the need for the self-assessment questionnaire to be tested in other independent

student groups. Chapter 3 describes the development and validation of this questionnaire. The results for tests of model fit, measurement invariance, and test-retest reliability suggest that the self-assessment questionnaire developed appeared to be an adequate instrument for measuring students' beliefs about the utility of self-assessment. Based on these findings, it is therefore important to examine whether the items of the questionnaire operate equivalently across different groups of students, for instance, age, gender, and experiences with engaging in self-assessment activities such as journal writing.

A second issue requiring further exploration refers to the use of identical instruments for self-, peer and tutor assessment. In the study discussed in Chapter 4, I examined students' self-assessment accuracy, as compared to the judgments by their teachers and peers. The outcomes of correlational analyses suggest that the partial overlap of the instruments may have produced, in part, the weak to moderate correlations between students' self-, peer and tutor assessment scores. Although this hypothesis was tested and rejected, another study employing identical instruments for self-, peer and tutor assessment should certainly be conducted to verify the findings reported in Chapter 4.

Given the range of students' aptitude and ability to cope with, and respond to, the task of assessing their own learning, the focus on individual students and their strengths and weaknesses should constitute the third line of research emerging from this thesis. This will help in better understanding the nature and operation of self-assessment in higher education. The gathering of detailed empirical evidence which may cast light on those characteristics and factors which could account for individual differences in student self-assessment skill is one key area for further research.

A fourth line of research emerging from this thesis comprises the investigation of whether student self-assessment skills can be improved through formal training in self-assessment. Feedback alone, as our study has demonstrated, is clearly not enough to affect change. Through a more structured and closely guided process, students may become better aware of, and value their existing capability for, self-assessment, and its potential for development and application. If students have better developed self-assessment skills,

it is likely that they will involve themselves in more effective learning and will thus become better metacognitive and self-reflective learners capable of critical evaluation of their own performance.

Finally, the study of the effects of reflection journal writing on student learning comes out from my thesis as a line of investigation. The study reported in Chapter 5 provided empirical support for the use of reflection journals to support student learning. Given that there is evidence of learning in student reflection journals, and if students do benefit from the activity of journal keeping, their academic performance is expected to improve. Results justify directing attention to further examine the effects of reflection journal writing on student academic achievement. Despite extensively stimulated by the literature, effective ways to improve students' academic performances through reflection journal writing have not been explored by empirical studies aimed on their design and test. Further research is required for that.

Knowledge accumulated from studies about self-assessment over the last decades provides some guidance for undertaking research initiatives to further develop students' self-assessment abilities. Findings of my studies revealed the potential effects of self-assessment on student learning, and avenues for exploring effective ways for exploiting the benefits of self-assessment on student academic achievement are therefore open for future research.

REFERENCES

- Alwis, W. A. M. (2007). Pedagogical Beliefs and Institutional Practices at Republic Polytechnic. *Keynote presented at the 2nd International Symposium on PBL: Reinventing PBL, Singapore.*
- American Psychological Association. (1997). *Learner-centered psychological principles*: Available online at: <http://www.apa.org/ed/lcp2/lcp14.html> (accessed 18 January 2008).
- Bain, J., Ballantyne, R., Packer, J., & Mills, C. (1999). Using journal writing to enhance student teachers' reflectivity during field experience placements. *Teachers and Teaching: Theory and Practice*, 5(1), 51-74.
- Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing Procedures for Implementing Peer Assessment in Large Classes Using an Action Research Process. *Assessment & Evaluation in Higher Education*, 27(5), 427-441.
- Blain, P., and William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-74.
- Boenink, A. D., Oderwald, A. K., Jonge, P. d., Tilburg, W. v., & Smal, J. A. (2004). Assessing student reflection in medical practice. The development of an observer-rated instrument: reliability, validity and initial experiences. *Medical Education*, 38, 368-377.
- Boud, D. (1986). *Implementing Student Self-assessment*: Kensington, N.S.W. : Higher Education Research and Development Society of Australasia.
- Boud, D. (1989). The role of self assessment in student grading. *Assessment & Evaluation in Higher Education*, 14(1), 20-30.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education*, 18(5), 529-549.
- Boud, D., Keogh, R., & Walker, D. (1985). *Reflection: Turning Experience into Learning*. London: Kogan Page.
- Brace, N., Kemp, R., & Snelgar, R. (2006). *SPSS for Psychologists*. London: Lawrence Erlbaum Associates.
- Britton, B. K., & Giilgoz, S. (1991). Using Kintsch's Computational Model to Improve Instructional Text: Effects of Repairing Inference Calls on Recall and Cognitive Structures. *Journal of Educational Psychology*, 83(3), 329-345.
- Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.
- Byrne, B. M. (2001). *Structural Equation Modelling With AMOS: Basic Concepts, Applications, and Programming*. London: Lawrence Erlbaum Associates.

- Cassidy, S. (2007). Assessing "Inexperienced" Students' Ability to Self-Assess: Exploring Links with Learning Style and Academic Personal Control. *Assessment and Evaluation in Higher Education, 32*(3), 313-330.
- Chirema, K. D. (2007). The use of reflective journals in the promotion of reflection and learning in post-registration nursing students. *Nurse Education Today, 27*(3), 192-202.
- Chua, J. C., & Schmidt, H. G. (2007). Are Teachers Able to Judge the Competency of Students Based on Classroom Observations? *Poster presented at the 2nd International Symposium on PBL: Reinventing PBL, Singapore.*
- Cohen, J. (1969). *Statistical Power Analysis for the Behavioural Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J., and Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences*. (Second ed.). Hillsdale, N.J. : Lawrence Erlbaum Associates.
- Crepe, P. (2005). Should student learning journals be assessed? *Assessment & Evaluation in Higher Education, 30*(3), 287-296.
- Dart, B. C., Boulton-Lewis, G. M., Brlwolee, J. M., & McCrindle, A. R. (1998). Change in knowledge of learning and teaching through journal writing. *Research Papers in Education: Policy and Practice, 13*(3), 291-318.
- Dewey, J. (1991). *How we think*. Buffalo, N.Y: Prometheus Books (Originally published: Lexington, MA : D.C. Heath, 1910).
- Dochy, F., Segers, M., & Sluijsmans, D. M. A. (1999). The use of self-, peer and co-assessment in higher education: A review *Studies in Educational Evaluation, 24*(3), 331-350.
- Eva, K. W., Cunningham, J. P. W., Reiter, H. I., Keane, D. R., & Norman, G. R. (2004). How Can I Know What I Don't Know? Poor Self Assessment in a Well-Defined Domain. *Advances in Health Sciences Education, 9*(3), 211-224.
- Falchikov, N. (2005). Improving Assessment Through Student Involvement: Practical Solutions For Aiding Learning in Higher and Further Education: (London, Routledge).
- Falchikov, N., and Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research, 59*(4), 395-430 Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287-322.
- Fitzgerald, J. T., White, C. B., and Gruppen, L. D. (2003). A longitudinal study of self-assessment accuracy *Medical Education, 37*(7), 645– 649.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906-911.

- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Gibbs, G. (1995). *Assessing Student Centred Courses*. Oxford: Oxford Centre for Staff Development
- Gleaves, A., Walker, C., & Grey, J. (2008a). Using digital and paper diaries for assessment and learning purposes in higher education: a case of critical reflection or constrained compliance?. *Assessment & Evaluation in Higher Education*, 33(3), 219-231.
- Gleaves, A., Walker, C., & Grey, J. (2008b). Using digital and paper diaries for learning and assessment purposes in higher education: a comparative study of feasibility and reliability. *Assessment & Evaluation in Higher Education*, 32(6), 631-643.
- Gordon, M. J. (1992). Self-assessment programs and their implications for health professions training. *Academic Medicine*, 67, 672-679.
- Grant, A., Kinnersley, P., Metcalf, E., Pill, R., & Houston, H. (2006). Students' views of reflective learning techniques: an efficacy study at a UK medical school. *Medical Education*, 40(4), 379-388.
- Haigh, M. (2001). Constructing Gaia: using journals to foster reflective learning. *Journal of Geography*
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems, in: R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural Equation Modelling: Present and Future- A Festschrift in honour of Karl Jöreskog*. Lincolnwood, IL: Scientific Software International, Inc. .
- Hays, W. L. (1988). *Statistics* ((4th ed.) ed.). Fort Worth, TX: Holt, Rinehart and Winston Inc.
- Herrero, A. H. (2007). Journals: A tool to improve students' writing skills [Electronic Version]. *Actualidades Investigativas en Educación*, 7, 1-37. Retrieved 5 Aug 2008.
- Hu, L.-t., & Bentler, P. M. (1999). Cut-off Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modelling*, 6(1), 1-55.
- Hubbs, D. L., & Brand, C. F. (2005). The Paper Mirror: Understanding Reflective Journaling. *Journal of Experiential Education*, 28(1), 60-71. *in Higher Education*, 25(2), 167-189.
- Kelson, A. C. M., & Distlehorst, L. H. (2000). Groups in problem-based learning (PBL): Essential elements in theory and practice. In D. H. Evenson & C. E. Hmelo (Eds.), *Problem-Based Learning: A Research Perspective on Learning Interactions*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kember, D., Jones, A., Loke, A., Mckay, J., Sinclair, K., Tse, H., et al. (1999). Determining the level of reflective thinking from students' written

- journals using a coding scheme based on the work of Mezirow. *International journal of lifelong learning*, 18(1), 18-30.
- Kerka, S. (1996). *Journal Writing and Adult Learning*: Available online at: <http://www.cete.org/acve/docgen.asp?tbl=digests&ID=27> (accessed 26 February 2006).
- Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel, .
- Kolb, D. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs: Prentice Hall.
- Kraayenoord, C. E. v., & Paris, S. G. (1997). Australian students' self-appraisal of their work samples and academic progress. *Elementary School Journal*, 97(5), 523-537.
- Langer, A. M. (2002). Reflecting on Practice: using learning journals in higher and continuing education. *Teaching in Higher Education*, 7(3), 337-351.
- Lew, M. D. N., & Schmidt, H. G. (2006). *Reflection upon learning between theory and practice: A focus-group study of tutors' and students' perceptions*. Erasmus University Rotterdam The Netherlands.
- Lew, M. D. N., & Schmidt, H. G. (2007a). Reflecting on Practice: the use of journals at a problem-based learning school in Singapore. *Poster presented at the 2nd International Symposium on PBL: Reinventing PBL, Singapore*.
- Lew, M. D. N., and Schmidt, H. G. (2007b). *Measuring students' beliefs about self-assessment*. Erasmus University Rotterdam The Netherlands.
- Lopez, R., & Kossack., S. (2007). Effects of Recurring Use of Self-Assessment in University Courses. *The International Journal of Learning*, 14(4), 203-216.
- Maclellan, E. (2001). Assessment for Learning: the differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, 26(4), 307-318.
- Maguire, S., Evans, S. E., & Dyas, L. (2001). Approaches to Learning: a study of first-year geography undergraduates *Journal of Geography in Higher Education*, 25(1), 95-107.
- Marton, F., & Säljö, R. (1984). Approaches to learning. In F. Marton, D. Hounsell & N. Entwistle (Eds.), *The Experience of Learning*. Edinburgh: Scottish Academic Press.
- McCombs, B. L., & Whistler, J. S. (1989). The role of affective variables in autonomous learning. *Educational Psychologist*, 24(3), 277-306.
- McCrinkle, A. R., & Christensen, C. A. (1995). The impact of learning journals on metacognitive and cognitive processes and learning performance. *Learning and Instruction*, 5, 167-185.

- McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Training International*, 32(4), 302-313.
- Meyers, C., & Thomas, J. B. (1993). *Promoting active learning* (San Francisco, CA, Jossey-Bass).
- Mezirow, J., & Associates. (1990). *Fostering Critical Reflection in Adulthood: A Guide to Transformative and Emancipatory Learning*. San Francisco: Jossey-Bass.
- Mezirow, J. (1991). *Transformative Dimensions of Adult Learning*. San Francisco: Jossey-Bass.
- Microsoft. (2008a). Microsoft Office Online: Grammar and writing style options. Available online at: <http://office.microsoft.com/en-us/word/HP051901251033.aspx> Retrieved 5 August, 2008
- Microsoft. (2008b). Microsoft Office Online: Readability scores. Available online at: <http://office.microsoft.com/en-us/word/HP051863181033.aspx?pid=CH060830131033> Retrieved 5 August, 2008
- Miller, G. A. (2006). WordNet: A lexical database for the English Language (Version 3.0). NJ: Princeton University.
- Mires, G. J., Ben-David, Friedman, M., Preece, P. E., & Smith, B. (2001). Educational Benefits of Student Self-Marking of Short-Answer Questions. *Medical Teacher*, 23(5), 462-466.
- Mok, M. M. C., Lung, C. L., Cheng, D. P. W., Cheung, R. H. P., and Ng, M. L. (2006). Self-assessment in higher education: experience in using a metacognitive approach in five case studies. *Assessment and Evaluation in Higher Education*
- Moon, J. A. (1999a). *Learning journals : a handbook for academics, students and professional development*. London, New York: Kogan Page.
- Moon, J. A. (1999b). *Learning journals : a handbook for reflective practice and professional development*. USA and Canada: Routledge, Taylor & Francis Group.
- Morrison, K. (1996). Developing reflective practice in higher degree students through a learning journal. *Studies in Higher Education*, 21(3), 317-332.
- Nicol, D., and Milligan, C. (2006). Rethinking technology-supported assessment in terms of the seven principles of good feedback practice, in: Bryan, C. and Clegg, K. (Eds). In *Innovative Assessment in Higher Education*: (London, Routledge).
- O'Connell, T., & Dymont, J. (2006). Reflections on using journals in higher education: a focus group discussion with faculty. *Assessment & Evaluation in Higher Education*, 31(6), 671-691.
- O'Rourke, R. (1998). The Learning Journal: from chaos to coherence. *Assessment & Evaluation in Higher Education*, 23(4), 403-413.

- Orsmond, P., Merry, S., & Reiling, K. (1997a). Students and tutors perceptions of a good essay. *Research in Education*, 58, 81-84.
- Orsmond, P., Merry, S., & Reiling, K. (1997b). A study in self-assessment: Tutor and students' perceptions of performance criteria. *Assessment & Evaluation in Higher Education*, 22(4), 357-369.
- Papinczak, T., Young, L., & Groves, M. (2007). Peer assessment in problem-based learning: A qualitative study. *Advances in Health Sciences Education*, 12(2), 169-186.
- Paris, S. G., & Cunningham, A. E. (1996). Children becoming students. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology*. In (pp. 117-147). New York: Simon & Schuster Macmillan.
- Paris, S. G., & Paris, A. H. (2001). Classroom Applications of Research on Self-Regulated Learning. *Educational Psychologist*, 36(2), 89-101.
- Paterson, B. L. (1995). Developing and maintaining reflection in clinical journals. *Nurse Education Today*, 15(3), 211-220.
- Pond, K., Ul-Haq, R., & Wade, W. (1995). Peer review: A precursor to peer assessment. *Innovations in Education and Training International*, 32(4), 314-323.
- Race, P. (1998). Practical pointers on peer assessment. In S. Brown (Ed.), *Peer assessment in practice*. Birmingham: SEDA publications.
- Sambell, K., McDowell, L., & Brown, S. (1997). 'But is it fair?': an exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23(4), 349-371.
- Schmidt, H. G., Vermeulen, L., & van der Molen, H. T. (2006). Longterm effects of problem-based learning: a comparison of competencies acquired by graduates of a problem-based and a conventional medical school. *Medical Education*, 40, 562-567.
- Schön, D. A. (1983). *Reflective Practitioner: how professionals think in action*. London: Temple Smith.
- Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35(4), 453-472.
- Searby, M., and Ewers, T. (1997). An evaluation of the use of peer assessment in higher education: A case of study in the school of music. *Assessment and Evaluation in Higher Education*, 22(4), 371-383.
- Segers, M., & Dochy, F. (2001). New assessment forms in problem-based learning: the value-added of the students' perspective. *Studies in Higher Education*, 26(3), 327-343.
- Selke, C. L., Petersen, B. T., & Nahrgang, C. L. (1986). Journal writing in mathematics. In A. Young & T. Fulwiler (Eds.), *Writing Across the Disciplines* (pp. 192-207). Upper Montclair, NJ: Boynton/Cook.

- Sluijsmans, D. M. A., Moerkerke, G., MerrKnboer, J. J. G. v., & Dochy, F. J. R. C. (2001). Peer Assessment in Problem based Learning. *Studies in Educational Evaluation*, 27(2), 153-173.
- SPSS, I. (2006). *SPSS Text Analysis for Surveys™ 2.0 User Guide*. Chicago, IL.
- Stefani, L. A. J. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Educational Evaluation*, 19(1), 60-75.
- Struyven, K., Dochy, F., & Janssens, S. (2002). *Student perceptions about assessment in higher education: a review*. Available online at: <http://www.leeds.ac.uk/educol/documents/00002255.htm> (accessed on 3 November 2006).
- Sullivan, K., and Hall, C. (1997). Introducing Students to Self-assessment. *Assessment and Evaluation in Higher Education*, 22(3), 289- 305.
- Sumsion, J., & Fleet, A. (1996). Reflection: can we assess it? Should we assess it? *Assessment & Evaluation in Higher Education*, 21(2), 121-130.
- Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: towards transparency for students and for tutor. *Assessment & Evaluation in Higher Education*, 26(6), 605-614.
- Thomas, J. (1999). *The Challenges of Developing Self Assessment Practices in Different Disciplines and Institutions: Educational development or educational research?* : Available online at: http://www.eurodl.org/materials/contrib/1999/assessment_issue/thomas.html (accessed on 3 November 2006).
- Thompson, G., Pilgrim, A., and Oliver, K. (2005). Self-assessment and Reflective Learning for First-year University Geography Students: A Simple Guide or Simply Misguided? . *Journal of Geography in Higher Education*, 29(3), 403- 420.
- Thorpe, K. (2004). Reflective learning journals: from concept to practice. *Reflective Practice*, 5(3), 327-343.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.
- Van Daalen, M. (1999). Test Usefulness in Alternative Assessment. *Dialog on Language Instruction*, 13(1&2), 1-26.
- Vockell, E. (2004). *Educational Psychology: A Practical Approach*: Available online at: <http://education.calumet.purdue.edu/vockell/EdPsyBook/> (accessed on 14 February 2008).
- Voss, M. M. (1988). The light at the end of the journal: a teacher learns about learning. *Language Arts*, 65(7), 669-674.
- Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring Self-assessment: Current State of the Art. *Advances in Health Sciences Education*, 7(1), 63-80.

- Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 315-327). New York: Macmillan.
- Wong, F. K. Y., Kember, D., Chung, L. Y. F., & Yan, L. (1995). Assessing the level of student reflection from reflective journals. *Journal of Advanced Nursing*, 22, 48-57.
- Wong, F. K. Y., Kember, D., Chung, L. Y. F., & Yan, L. (1995). Assessing the level of student reflection from reflective journals. *Journal of Advanced Nursing*, 22, 48-57.
- Woodward, H. (1998). Reflective Journals and Portfolios: learning through assessment. *Assessment and Evaluation in higher education*, 23(4), 415-423.
- Yinger, R. (1985). Journal writing as a learning tool. *Volta Review*, 87(5), 21-33.
- Zimmerman, B. J. (2000). Attaining Self-Regulation: a social cognitive perspective. In M. Boekaerts, P. Pintrich & M. Zeodmer (Eds.), *Handbook of Self-Regulation*: Academic Press.

Appendix A: Self- and Peer assessment statements

Statements of Self-assessment

1. Listening to and valuing what others had to say.
2. Saying what I wanted to say clearly.
3. Encouraging others to share what they thought.
4. Pointing out any disagreements or contradictions of ideas that had been raised.
5. Pointing out any agreements or connections between ideas that had been raised.
6. Suggesting a hypothesis or a possible solution built on the ideas of the group.
7. Making reference to something I read to support or refute an idea.
8. Asking a question that warrants further investigation.

Statements of Peer assessment

1. The team member was cooperative.
2. This team member completed the tasks assigned by our team.
3. This team member did more than what was expected.
4. This team member contributed useful ideas.

Appendix B: Questions for focus-groups with tutors

1. Let us now discuss the elements of the daily assessment, starting with the self- and peer assessment, the reflection journal, followed by the tutor judgment (observations followed by tutor feedback).

Self-assessment

- a. What do you think is the purpose of the self-assessment?
- b. What can your students benefit from the self-assessment?
- c. How has the self-assessment helped you in your grading? (Probes about trends and reliability of students' self assessments)

Peer assessment

- a. What do you think is the purpose of the peer assessment?
- b. What can your students benefit from the peer assessment?
- c. How has the peer assessment helped you in your grading? (Probes about trends and reliability of students' peer assessments)

Reflection journal

- a. What do you think is the purpose of the reflection journal?
- b. How can your students benefit from writing the reflection journal?
- c. What type of reflection journal questions do you usually ask your students? (Probes about their preference of getting students to report factual information or reflect on their learning and development)
- d. How has the reflection journal helped you in your grading?

Tutor judgment

- a. How has observing the students' performance helped you in your grading?
- b. What do you think is the purpose of the tutor feedback? (Probes about whether observations are used formatively or summatively or both)
- c. Do you give feedback to your students based on your observations of how they performed during the day? If yes,
 - i. Why do you give feedback?
 - ii. What type of feedback do you give?
 - iii. How do you think your students can benefit from the feedback?

Let us now talk about the daily grades in general.

Describe briefly the process how you derive daily grades (Probes about which assessment tools are taken into consideration, and which has the greatest influence on daily grade).

Appendix C: Questions for focus-groups with students

Let us now discuss the elements of the daily assessment, starting with the self- and peer assessment, the reflection journal, followed by the tutor judgment (observations by your tutors and tutor feedback).

Self-assessment

- a. What do you think is the purpose of the self-assessment?
- b. In what ways is the self-assessment useful?
- c. How do you feel about assessing your own performance? (Probes about reliability of self-ratings, daily rigor)
- d. How do you feel about the self-assessment as a criterion for grading and why?

Peer assessment

- a. What do you think is the purpose of the peer assessment?
- b. In what ways is the peer assessment useful?
- c. How do you feel about being assessed by your peers? (Probes about benefits and problems; reliability of peer assessments)
- d. How do you feel about assessing your peers? (Probes about benefits and problems; reliability of peer assessments)
- e. How do you feel about the peer assessments as a criterion for grading and why?

Reflection journal

- a. What do you think is the purpose of the reflection journal?
- b. What type of reflection journal questions do you prefer: questions that ask you to detail content knowledge, or questions asked you to reflect on your learning and development? Why?
- c. In what ways has doing the reflection journal helped you in your learning?
- d. Do you write about other “things” other than your response to the reflection journal question in your in your reflection journal? If yes, give examples.
- e. How do you feel about the reflection journal as a criterion for grading and why?

Tutor judgment

- a. Do you adopt strategic behaviours for different tutors so as to improve your grade?
- b. What type of feedback do your tutors give? Do you find it useful? Why?
- c. Which assessment tool do you think is the most/least useful and why?

Appendix D: Self-assessment section of the Self-Assessment Questionnaire

Table 1. Items belonging to the self-assessment section of the Self-Assessment Questionnaire

Construct		Item
The self-assessment enables me to make an appraisal of my learning.	1	Doing the self-assessment enables me to judge my performance better.
	2	The self-assessment enables me to improve on my learning in areas that I'm not so good at.
	3	I become better aware about my learning through doing the self-assessment.
	4	The self-assessment helps me to assess my strengths and weaknesses accurately.
Assessing my own performance is more of a habitual action than to improve on my learning.	5	Doing the self-assessment everyday is too frequent.
	6	Doing the self-assessment is a waste of time.
	7	I do the self-assessment without thinking how the statements are related to my performance during the day.
The self-assessment enables me to manage my tutor's impressions of how I performed.	8	The self-assessment is mainly useful in managing the tutor's impression of my performance.
	9	My tutor looks at my self-assessment when he/she grades.
	10	I assess myself in order for the tutor to grade me.

Note. This table is adapted from the full version of the validated questionnaire (Lew and Schmidt 2007b). The constructs underlying the questionnaire are described by sentences rather than phrases. It is perhaps more informative to describe the underlying construct in the form of a statement.

Appendix E: Grammar, Readability and Coherence tests

Grammar test

The grammar check options in Microsoft Office Word and what they detect are described in the following table.

Table 1. Grammar check options: descriptions and examples

Grammar check option	Description	Example
Capitalization	Capitalization problems such as proper nouns or titles that precede proper nouns. Also detects the overuse of capitalization.	"Mr. jones" instead of "Mr. Jones"; "miss Helen" instead of "Miss Helen"
Fragments and Run-ons	Sentence fragments and run-on sentences.	"That we had a good idea on explaining the concept of genetic hereditary."
Misused words	Incorrect usage of adjectives and adverbs, comparatives and superlatives, units of measure, conjunctions, prepositions, and pronouns.	"like" as a conjunction; "nor" versus "or"; "what" versus "which"; "who" versus "whom"
Negation	Use of multiple negatives.	"I don't can understand a single word the tutor said" instead of "I can't understand a single word the tutor said"
Noun phrases	Incorrect noun phrases; a/an misuse; number agreement problems in noun phrases.	"five team mate" instead of "five team mates"
Possessives and plurals	Use of a possessive in place of a plural, and vice versa. Also detects omitted apostrophes in possessives.	"A lot of us disagrees with Ben's explanation" instead of "A lot of us disagree with Ben's explanation"
Punctuation	Incorrect punctuation, including commas, colons, end-of-sentence punctuation, punctuation in quotations, multiple spaces between words, or a semicolon used in place of a comma or colon.	"I enjoy coming to school, and interacting with my class mates." instead of "I enjoy coming to school and interacting with my class mates."
Questions	Non-standard questions	"I asked my team mate if he could show me how to plot an Excel graph?"
Relative clauses	Incorrect use of relative pronouns and punctuation, including	"who" used in place of "which" to refer to things,

Grammar check option	Description	Example
		"which" used in place of "who" to refer to people, unnecessary use of "that" with "whatever" and "whichever" or "that's" used in place of "whose"
Subject-verb agreement	Disagreement between the subject and its verb, subject-complement agreement, and subject-verb agreement with pronouns and quantifiers.	"All of the teams have presented their findings." instead of "All of the teams have presented their findings."
Verb phrases	Incorrect verb phrases; incorrect verb tenses; transitive verbs used as intransitive verbs.	"I has been reading about DNA" instead of "I have been reading about DNA"

Note: Information adapted from Microsoft (2008a).

Readability tests

The two readability tests used in the present studies are the Flesch Reading Ease and the Flesch-Kincaid Grade Level. Each readability test bases its rating on the average number of syllables per word and words per sentence (Flesch, 1948; Kincaid et al., 1975; Microsoft, 2008b).

The Flesch Reading Ease score is computed by means of the following formula:

$$206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW})$$

The Flesch-Kincaid Grade Level score is computed by means of the following formula:

$$(0.39 \times \text{ASL}) + (11.8 \times \text{ASW}) - 15.59$$

Where:

ASL = average sentence length (the number of words divided by the number of sentences)

ASW = average number of syllables per word (the number of syllables divided by the number of words)

Coherence test

In the present study, coherence is represented by means of an index, ranging from 0-1 and which is computed using the given formula:

$$CI = \frac{P}{S - 1}$$

Where:

CI = coherence index

P = total number of pairs of identical words or synonyms between a segment and the segment which immediately follows it

S = total number of segments in a given body of text

Note that a segment is created through segregating the text by means of conjunctions, which are essentially parts of speech that connect two or more words or phrases. Examples of conjunctions include "and", "or", "but", "yet", "because" etc. Text is also segregated by means of punctuation marks like full stop (.), comma (,) and semicolon (;).

As illustration, we used an example of a student's journal response in demonstrating how the Microsoft macro we developed in determining textual coherence works.

Sample student journal response:

"I can contribute better to my team's learning by expressing my ideas. It is good that I share my ideas with them. Another thing is I need to maintain a good relationship with my team mates. When I do not know how to make the computer program I need to ask them. I must thank my team mates for helping me to understand the program. I can also help by doing some research on the internet."

First, the macro divides the text into shorter segments through segregation by conjunctions.

Segment 1: I can contribute better to my team's learning by expressing my ideas

Segment 2: It is good that I share my ideas with them

Segment 3: Another thing is I need to maintain a good relationship with my team mates

Segment 4: When I do not know how to make the computer program I need to ask them

Segment 5: I must thank my team mates for helping me to understand the program

Segment 6: I can also help by doing some research on the internet

Second, the macro identifies identical words in a given segment by comparing with the segment that immediately follows it. If there is more than one instance The program generates the output as follows:

Number of match between segment 1 and 2: 1 (ideas: ideas; my: my)

Number of match between segment 2 and 3: 1 (good: good; my: my)

Number of match between segment 4 and 5: 1 (program: program)

Total number of matches = 3

The macro then computed the coherence index of the student journal response.

$$CI = \frac{3}{7-1} = 0.50$$

Unlike sophisticated methods of measuring coherence such as latent semantic analysis and topical structure analysis which use sophisticated analysis techniques, the test of coherence described here gives a crude approximation of how well the meanings and sequences of ideas relate to each other in a given text.

CURRICULUM VITAE

Duan Ning Magdeleine Lew was born on July 12, 1980 in Singapore. She completed her secondary education at the Nanyang Girls' High School in Singapore. In June 2003, she graduated with a Bachelor's degree in Chemical Engineering from the National University of Singapore. Upon graduation, she joined Republic Polytechnic Singapore as an academic staff with the School of Applied Science, where she involved herself mainly in the development of Biopharmaceutical Technology modules. In April 2006, she joined the Polytechnic's Office of Academic Services, and focused mainly on coordinating modules and activities for first-year students. She was also involved in developing policies and academic web pages for staff. Magdeleine was awarded the Masters of Science (Biomedical Engineering) from the Nanyang Technological University, Singapore in June 2006. From July 2006 onwards, she combined her work at Republic Polytechnic with part-time Doctor of Philosophy program offered by the Department of Psychology at the University of Erasmus Rotterdam. Her research concerned students' self-assessment in higher education and resulted in the present dissertation. Since November 2008, Magdeleine joined the Polytechnic's Centre of Educational Development. She is now heavily involved in staff developmental training, and providing consultancy services to other departments in terms of facilitation and problem crafting.

PUBLICATION

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (in press) Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education*

PRESENTATIONS

Lew, M. D. N., & Schmidt, H. G. (2007). Reflecting on Practice: the use of journals at a problem-based learning school in Singapore. *Poster presented at the 1st International Symposium on PBL: Reinventing PBL, Singapore.*

Lew, M.D.N. & Schmidt, H.G. (2007). Online reflection journals: Learning through assessment. In *ICT: Paper presented at the Ascilite 2007: Providing choices for learners and learning, Singapore.*

Lew, M. D. N., W.A.M, A., & Schmidt, H. G. (2008). Measuring students' beliefs about self-assessment. *Paper presented at the 15th International Conference on Learning, Chicago USA.*

Lew, M. D. N., W.A.M, A., & Schmidt, H. G. (2008). Peer Assessment in Problem-based Learning: Students' Views. *Paper presented at the 34th IAEA Annual Conference, Cambridge, UK.*

Lew, M. D. N., W.A.M, A., & Schmidt, H. G. (2008). Self-Assessment in a One-Day, One-Problem Problem Based Learning Curriculum: Students' Views. *Paper presented at the TLHE 2008: International Conference on Teaching and Learning in Higher Education, National University of Singapore.*

Lew, M. D. N., W.A.M, A., & Schmidt, H. G. (2009). A Longitudinal study of Students' Self-Assessment Accuracy. *presented at the 2nd International Symposium on PBL: What are we learning about learning?, Singapore*