ACTA DIDACTICA NAPOCENSIA
Volume 3, Number 1, 2010

A RUBRIC TO SELF-ASSESS AND PEER-ASSESS MATHEMATICAL PROBLEM SOLVING TASKS OF COLLEGE STUDENTS

Gunawardena Egodawatte

Abstract. Student involvement in their own assessment can add reflection and metacognition to the learning process. Based on this idea, an assessment instrument was developed to self-assess college students' mathematical problem solving tasks. The main objective of this exercise is to improve student learning. The assessment instrument contains three items: student opinion questionnaire, an analytic evaluation rubric, and a self assessment schedule. A peer assessment component was also included in the assessment process to gain more accuracy. The rubric will serve three main purposes. First, students can use the rubric as a learning tool. Second, it will serve as a teaching tool for teachers. Third, it can be used as an assessment tool. However, the whole purpose is to improve student learning rather than providing them with a letter grade.

Keywords: Assessment, self assessment, evaluation, rubrics, mathematical problem solving.

Introduction

In higher education, the focus in recent times has been on issues relating to widening participation and student employability (Cassidy, 2007). As a consequence, higher education as a whole is not only serving record numbers of students but is required to make explicit employability skills which include a broad range of skills or competencies that are not limited to technical skills such as subject specialisms but are transferable or non-technical skills. The need for students to develop as independent learners is both fundamental to academic success in higher education and essential to subsequent professional success. One defining characteristic of independent learners is there ability to self-assess. This involves a high level of self-awareness and the ability to monitor one's own learning and performance.

According to the Conference Board of Canada (1992), there are three categories of critical skills required for the Canadian workforce. These are problem solving skills, teamwork skills, and other soft skills such as responsibility, accountability, self management, recognition of and respect for people's diversity and individual differences. These skills are meant to be promoted in all learning situations in college classrooms. Therefore, it is important to teach as well as to assess these skills in a student-centred environment. In order to enlarge the educational congruence, students should be given responsibility in the assessment practices. Using self assessment is one way of assessing students and improving their learning. It is recommended in education that students should be able to reflect on their own behavior (self assessment) and that of other peers (peer assessment).

In this backdrop**,** this paper attempts to focus on constructing a rubric to assess college students' mathematical skills by self assessment and to use these results to improve student learning. The paper attempts to focus on four interrelated objectives. They are:

- To introduce a self assessment strategy for the appraisal and improvement of individual learning
- To enhance the possibility of this improvement through the supportive scrutiny of peers

Received 17 February 2010, accepted 16 March 2010, published 27 March 2010

- To provide an opportunity to reflect upon and review areas of one's own strengths and weaknesses
- To encourage individuals to identify ways in which they could improve aspects of their own learning.

Throughout this paper, the words "self assessment" and "self evaluation" are meant to have the same meaning and these words have been used interchangeably. Here, both terms mean "the involvement of students in identifying standards and/or criteria to apply to their work and making judgments about the extent to which they have met these criteria and standards (Boud, 2003, p.12)".

The purpose of self assessment

Most often, there is a component of assessment or evaluation which involves with many kinds of learning situations. According to Nitko (1996), assessment is a broad term defined as a process for obtaining information that is used for making decisions about students, curricula and programs, and the educational policy. According to Orsmond et al. (2000), there are three primary reasons as to why assessment procedures are carried out. They are to facilitate student learning, to enable students to become reflective practitioners, and to provide students with formal accountability and accreditation of knowledge. Therefore, improve learning is one important aspect of assessment.

Among other assessment methods, self assessment is a novel method of assessing students and improving their learning. Self assessment has three main purposes. It can be used as an educational outcome, a learning strategy or an assessment tool. It also relates to the intrapersonal intelligence proposed by Garner (Manning, 1993). This intelligence type relates to the ability to step back and reflect to understand one's feelings, actions, and abilities. The idea is to reflect on and reflect in one's own actions to identify the weaknesses and strengths. According to Boud (2003), all assessment whether conducted by teachers or by learners has two functions. The first function is the development of knowledge and an appreciation of the appropriate standards and criteria. The second function is the capacity to make judgments about the work involved and to see whether the work does or does not meet the standards. In this paper, the focus of assessment is to improve student learning. Self assessment engages students in evaluating their progress, developing their communicative skills, and increasing their mathematical vocabulary (Stallings & Tascione, 1996).

Boud (2003) discussed the uses of self assessment in eight major categories. They are: for individual self-monitoring and checking progress, for promoting how-to-learn skills, for diagnosis and remediation, as a substitute for other forms of assessment, to improve professional or academic practice, to consolidate learning over a wide range of contexts, to review achievement as a prelude to recognize prior learning, and for self knowledge and self understanding. Boud further stated that all acts of assessment, whether by teachers, subject matter experts, peers, or the individual learner, involve establishing criteria and judging work in the light of these criteria. Judgment based on criteria is an emphasis in this definition.

Watson (2006) said that innovative informal assessment practices relate to significant learning gains. Watson elaborates five domains which make a difference to achievement. Goal orientation indicates that assessment should focus on processes rather than the product. Self-perception is the feedback that needs to develop a personal sense of worth and progress. Assessment by students will give students the responsibility and this improves their motivation and achievement. Assessment is a moment in a holistic learning process, incorporating a reflective, meta-cognitive mode. Therefore, it should be linked to theories of learning. Hence, cases of practice emerge as key to effective formative assessment. Self and peer assessment are important skills in order to work on complex problems is widely acknowledged in education (Birenbaum & Dochy, 1996; Boud, 2003; Sambell & McDowell, 1998). They further stated that it is important to plead for an assessment system that requires students to use higher-order thinking skills to solve and analyze problems instead of memorizing facts and solving well structured, decontextualized problems.

Students should be able to reflect on their own behavior (self assessment) and that of other peers (peer assessment**).** By allowing students to take responsibility of their own learning through self assessment, teachers are empowering their students partly drawing out the authority from the teachers. In a supporting argument, Hendry (1996) stated that by involving students in self-assessment activities which provide students with the opportunity to develop metacognitive and more general learning skills, tutors are empowering students. In terms of student empowerment, Stefani (1994) contends that many lecturers/tutors express great fear of handing any of the power of assessment over to students. The fear, according to Stefani, generally stems from the possibility that the student marks will differ significantly from lecturer marks. To counteract this fear, it can be argued that introducing students to self and peer assessment early in their academic career and using the mark summatively as well as formatively will engender a sense of responsibility in students by the time that the grading and ranking of them becomes a crucial matter.

A number of important themes emanated from the above discussion. Learners' ability to identify areas of performance that require a greater degree of improvement would lend greater learner efficiency to self-directed learning. Student ownership of the assessment criteria is fundamentally important. Student ownership of assessment procedures comes through the empowerment of the students. This ownership by students has to emanate from the generation of criteria by themselves. Ownership involves elements of student participation in the design of the process rather than tutor imposition of the design. The terminal feedback of any assessment procedure to students is a better form of allowing students to influence the future design of the scheme.

The previous discussion also shows the importance of training students before any self assessment process thereby alleviating the fears of the students and the teachers. Moreover, self assessment provides students with metacognitive and judgmental skills together with opportunities to improve their learning. To gain assessment skills and learning skills, students should provide every possible opportunity to self assess. Hence, using the guidelines of a self assessment instrument such as a scoring rubric facilitates both evaluation and self learning and provides a blueprint for both the grade-motivated and the knowledge-motivated student. Thus, the importance of rubrics in the self assessment process is discussed in the next section.

Rubrics-based assessment

Wiggins (1998) says the word rubrics derives from *ruber*, the Latin word for red. In medieval times a rubric was a set of instructions or a commentary attached to a law or liturgical service and typically written in red. Thus, rubric came to mean something that authoritatively instructs people. However, in student assessment, a rubric is a set of scoring guidelines for evaluating students' work. Wiggins further says that typically a rubric contains a scale of possible points to be assigned in scoring on a continuum of quality. High numbers are usually assigned to the best performances. Moreover, a rubric provides descriptors for each level of performance to enable more reliable and unbiased scoring. Rubrics assessment is itself valid and reliable when proper descriptors are formulated because descriptors contain criteria which often refer to standards. The criteria are the conditions that any performance must meet to be successful. For example, the descriptor "effectively listen" may have two criteria as taking apt steps to comprehend and making the speaker feel heard. Arther & McTighe (2000) define a rubric as a set of general criteria used to evaluate a student's performance in a given outcome area. Rubrics consist of a fixed measurement scale and a list of criteria that describe the characteristics of products or performances for each score point. According to Nitko (1996), a scoring rubric is a coherent set of rules that we use to assess the quality of a student's performance. These rubrics may be in the form of a rating scale or a checklist.

Quinlan (2006) contrasts the differences between rating scales and checklists. She says that rubrics are better than checklists or performance lists as the latter items do not usually define clearly the criteria that the evaluator may have in mind. If teachers would like their students to reliably create quality material and thoroughly understand each scoring point, they may want to prefer a rubric than a checklist or a performance list. Quinlan cites five advantages of using precise sets of criteria that describe levels of performance or understanding in a scoring rubric. Rubrics provide students with: expectations about what will be assessed, information on the standards that need to be met, and

indications of where they are in relation to goals. They increase consistency in teacher ratings of performance, products, or understanding and provide teachers with data to support grades.

Rubrics for scoring constructed responses have two general categories: analytic scoring rubrics and holistic scoring rubrics. An analytic scoring rubric requires that the teacher identifies the important aspects of a good solution and then assign points to each aspect. On the other hand with a holistic scoring rubric, teachers must determine the overall quality of the constructed-response. Holistic rubrics place student work into predetermined categories that reflect the quality of the response. According to Nikto (1996), the holistic rubric could be extended to include comments and this creates another rubric type as annotated holistic rubrics.

The holistic rubric rate or score the product or the process as a whole without first scoring parts or components separately (Nikto, 1996). The analytic rubric rate or score separate parts or characteristics of the product or the process first and then sum these part scores to obtain a total score. In an annotated holistic rubric, raters use a holistic rating first and then they rate or describe a few characteristics that are strengths and weaknesses to support their holistic ratings. Among the three types, using an analytic scoring rubric is a more time-consuming task since the rater has to look for and separately rate each component of a performance. This level of detail is useful when the focus is on diagnosis or helping students to understand the expectations for each part of the performance. This may be especially useful for helping students to learn even though it is time-consuming. According to Arther & McTighe (2000), an analytic trait rubric divides a product or performance into essential traits or dimensions so that they can be judged separately. A separate score is provided for each trait. In general, the above principles apply to any learner in an assessment situation although it is important to look at how concepts in self assessment affect adult learners.

Use of rubrics in assessing adult learning

Adults bring different experiences to the classroom than high school students. Arthur (1995) believes that because self-evaluation is linked to self-direction, it is more appropriate than some other evaluation measures for adult learners who are considered to be mature, self-motivated individuals. Theories of learning for adult learners are different from non-adult learners. The theories for adult learning are called “andragogy” while the theories for non-adult learning are called “pedagogy”. Quinlan (2006) cited some learning differences between adult and non-adult learners. According to her, in terms of student self-concept, adults are independent learners who are able to self-direct their learning. In terms of student prior experiences, they often bring a wealth of life experiences to the learning situation. Adults often ready to learn the individual needs for his or her perceived role. This is cited as student readiness to learn. In terms of application of learning, what adults learn may sometimes need immediate applications in workplace situations. Therefore, adult students’ motivation is intrinsic as their learning has an immediate impact on their lives. Those so-called characteristics should be taken into account when preparing evaluation schemes for adult learners. Quinlan (1996) recommends eight critical successful ways of adult teaching. The important ones for self assessment are: value the experience of the learners, engage in reflection of the learners’ experiences, empower the students, assess each student as an individual, and encourage a willingness to make changes based on learning experiences.

One important idea from the above discussion is that adult students bring with them a variety of life experiences to the class. These experiences often include prior grading experiences, both positive and negative. Today, rubrics are widely used in many self assessment situations. Rubrics can be effectively employed to empower adult students, improve their learning, and assess them individually. Therefore, using rubrics with adult learners helps to address issues of fairness in grading by assessing each assignment as an individual work. Since a special emphasis has been given in this paper to assess adult mathematical problem solving tasks with the use of rubrics, how rubrics could be effectively used in assessing mathematical tasks will be discussed in the next section.

Use of rubrics in mathematics assessment

Rubrics can help teachers analyze and describe students’ responses to complex tasks and determine students’ levels of proficiency. In addition, rubrics give students more specific criteria detailing what is expected and what constitutes a complete response. Meier et al. (2006) reported a study in which

middle school teachers used rubrics on eighth grade students to score non-traditional mathematical tasks. The teachers used analytic scoring rubrics that outlined three categories for evaluation namely mathematical knowledge, strategic knowledge, and explanation. Mathematical knowledge addressed concerns related to mathematical accuracy and correctness of terminology, while strategic knowledge was related to the identification of important parts of the problem and discussion of the methods of the solution. The explanation category dealt with a description of what was done, and required a discussion of why it was done with a written explanation for any diagrams or tables.

Meier et al. (2006) further stated that under mathematical knowledge, students are expected to demonstrate their knowledge of mathematical concepts, principles, and procedures. This requires an understanding of relationships among problem elements and use of mathematical terminology and notation. It may also require students to recognize, when a procedure is appropriate, execute a procedure, verify results of a procedure, and generate or extend familiar procedures. Under strategic knowledge, students are expected to use models, diagrams, and symbols to represent and integrate concepts in addition to being systematic in their application of strategies. For some assessment tasks, students are expected to justify their answers. This justification requires an appropriate mode of communication (e.g. written, pictorial, graphical or algebraic methods) for expressing the integration of mathematical ideas, conjectures, and arguments. For other assessment tasks, students are required to describe their procedures or strategies. Ultimately, students are also expected to communicate their mathematical ideas in writing: symbolically, or visually, use mathematical vocabulary, notation, and structure to represent ideas, and describe relationships.

In a study conducted by Popham (1997), an analytic rubric was used to measure mathematical skills requiring students to complete three subtasks: averaging, graphing, and concluding. Some of the important features of this rubric were that each subtask had teachable evaluative criteria. Those criteria are applicable across a wide range of similar subtasks. Also, this rubric does not delineate the nuances of each evaluation criterion so that different people using the rubric would invariably score students' responses in an identical manner confirming instrument reliability. Lane (1993) provided a four component framework that provides guidelines in the construction of assessment tasks. The first component, cognitive processes and task demands, includes: understanding and representing problems, discussing mathematical relations, organizing information, using and discovering strategies and heuristics, using and discovering procedures, formulating conjectures, evaluating the reasonableness of answers, generating results, justifying answers or procedures, and communicating. The second component, mathematical content, includes content categories such as number and operations, measurement, estimation etc. Modes of representation is the third component which refers to the internal mental model. This includes problem solving constructs in the forms of written, pictorial, graphical, tabular forms and so on. Finally, task control is the component in which valid assessment tasks represent the ways in which knowledge and skills are used in "real world" contexts and stresses the need for embodying them. Some of those components are applicable to many types of mathematical tasks. The above sub categories are very lengthy and in my opinion, one has to choose a manageable number of criteria and performance tasks in order to have a handy instrument. A five-point rating scale of 0 to 4 would be a reasonable choice for an analytic scoring rubric.

The focus in this study is to construct an analytic scoring rubric to self assess and peer assess mathematical problem solving tasks. This rubric will serve three main purposes. First, students can use the rubric as a learning tool by identifying their own strengths and weaknesses. Second, the information gathered from this rubric will help the teacher to redesign future teaching and in that sense it will act as a teaching tool. Finally, this rubric can be used for student assessment purposes providing the teacher to use it as an assessment tool. My emphasis will be to assess the process of problem solving for better learning rather than evaluating the product.

Minimizing rating differences

Sluijsmans et al. (2001) revealed some personal differences among raters in their standards and their rating styles. They said that raters may differ in their severity or leniency. Some raters consistently tend to give high grades (lenient raters), while others consistently give low grades (severe rates). Sometimes raters differ in the extent to which they distribute grades on the score scale. Some raters

tend to distribute scores closely around the average while others will spread scores much more widely. In other words, some raters avoid giving extreme grades while others prefer to use them. Another effect is the halo effect. This is the tendency of human raters to base distinctive aspects of the rating on an overall impression created by one single dominating aspect. This may indicate that raters cannot differentiate among distinct aspects of one product or procedure. The significant effect refers to the fact that raters may have different opinions about the rating tasks. This problem is not so much related to the divergent views of an individual, but rather to the diverging opinions of groups of individuals.

Voss and post in Sluijsmans et al. (2001) argue that, in particular, in the assessment of "soft" or less "tangible" skills, objectivity is significantly decreased due to divergence of views among raters of different schools. Another rating error is caused by the evaluation policy. Judges differ in the way they employ criteria. Every assessor has his or her own evaluation policy. According to some of them, the performance must achieve a minimum qualifying level on a number of criteria. Other judges act conjunctively while the performance is excellent on one criterion and it is weak on the rest of the criteria. One could also judge compensatorily; poor showing on some criteria could be balanced by high performance on others. Peer assessment, the way it was conducted does not prevent rating errors like friendship marking, resulting in over-marking; collusive marking, resulting in a lack of differentiation within groups; decibel marking, where individual who dominate groups get the highest marks; and parasite marking, where students fail to contribute but benefit from group marks.

Miller (2003) contends that the value of peer and self-assessment is commonly diminished by scoring range restriction by the raters. He says that the more specific written feedback requests are better for peer evaluation using rubrics. Educators should consider the effects of criteria specificity and written feedback solicitation on rater behavior when designing the instruments. Miller cites from other researchers and says that peer and self assessment instruments commonly consist of 4-6 criteria for scoring, with the nature of the criteria depending on the assessment type whether it is an oral presentation, written paper or group work. Rating scales for these assessments usually involve the assignment of a numerical score for each item in a set of criteria, with 4-5 point scales being common. A problem sometimes reported with the use of peer and self-assessments has been the tendency of raters to assign a very narrow range of scores, usually at the high end of the rating scale which is called as the 'ceiling effect". This tendency is frequently related to issues concerning the use of students as assessors, including their lack of ability to discriminate levels of performance and their reluctance to judge their peers or themselves. However, Miller says, that the same scoring tendencies have been seen with faculty assessors as well. Scoring leniency and range restriction are significant problems for the utilization of peer and self-assessment. If the instruments do not allow for the discrimination of performance, they have little formative or summative value for the students being assessed. The problem with scoring leniency and range restriction may also be related to the scoring system used with these instruments. If the scoring criteria are too vague or difficult to understand or if the rating scale offers too few choices for scoring, then an accurate, fair judgment can be difficult to make, possibly causing rates to grade very highly, so as not to penalize anyone being assessed. According to Miller, one possible solution to this problem is to increase the specificity of the scoring criteria on the instrument, by increasing their number and targeting them at very discrete areas of student performance. This may reduce the problems associated with using only a few poorly defined criteria and give raters additional cues when assessing various components of student performance. Having discussed various concepts around self assessment and use of rubrics in this process, it is appropriate now to apply these ideas in a real classroom situation. The application selected in this paper is to construct a rubric for evaluating mathematical problem solving tasks of college students. This application will be explained next together with relevant background information.

Application

Selection of criteria and the connection between criteria and essential employability skills

In a community college classroom, students are mainly being prepared for future employment. As mentioned previously, the Canadian system of education expects community college students to practice certain skills in their classrooms that are essential in their future work. Henry Pollak, a noted industrial mathematician summarized the mathematical expectations for new employees in industry

(NCTM, 1989). Some important ideas in this summary are the ability to set up problems with appropriate operations, knowledge of a variety of techniques to approach and work on problems, understanding the underlying mathematical features of a problem, ability to see the applicability of mathematical ideas to common and complex problems, belief in the utility and value of mathematics, and the ability to work with others on problems.

Many community colleges in Canada formulate their course objectives compatible with the essential employability skills mandated by the Ministry of Training, Colleges, and Universities. These skills include effective communication skills, execution of mathematical operations accurately, apply systematic approaches to solve problems, use a variety of thinking skills, interact with groups, effective working relationships and so on. According to the Conference Board of Canada (1992), there are three categories of critical skills required for the Canadian workforce. The first category involves: problem solving in a variety of areas including mathematics and using the results, thinking critically, acting logically to evaluate situations, making decisions, and read, comprehend, and use written materials including graphs, charts, and displays. The second category includes responsibility, accountability, self management, and recognition of and respect for people's diversity and individual differences. The third teamwork skills category includes: understanding and working within the culture of the group, participating in the decision making process, respecting the thoughts and opinions of others in the group, seeking a team approach as appropriate and assuming a leadership role as appropriate.

In many college classroom situations, students often work in small groups. The group social skills of students are frequently being evaluated in assessment procedures as proposed in broader student outcome categories. However, hard skills such as individual problem solving skills are not frequently being assessed to improve student learning. Also, many students are unaware of their difficulties in the mathematical problem solving process until they get their evaluated test papers in the end. Therefore, students will better learn the techniques and skills that they are expected to learn when they are given a chance to self assess and when they are assessed by their peers. Evaluating cognitive skills is equally important as evaluating group social skills. Therefore, in this paper, I wish to address this issue by constructing a rubric for students and their peers to assess each others' problem solving skills.

Schoenfeld (2002) stresses the importance of assessing domain-specific problem solving skills as same as other workplace skills. According to Schoenfeld, not only workplace-oriented skills are essential to evaluate in apprenticeship programs but most importantly other skills such as problem solving, metacognition, and non-domain specific task skills are also important aspects to evaluate as skills that are ever changing with rapid changes taking place in today's society. Today, most proposed school-to-work programs are oriented toward skills. These programs may proceed by examining the workplace, identifying productive skills, and teaching them directly. Or, they may take a more "contextual" approach, with suggestions that students should engage in apprenticeships, that curricula should be designed to reflect workplace demands, or both. According to Schoenfeld, such approaches are doomed to fail. There is, of course, a pragmatic reason.

> "The skills set is a moving target in that skills learned today will be obsolete tomorrow, and new skills will be needed. More important, there is a deep theoretical reason. The past quarter century of research in mathematics education has shown that skills are but one component of mathematical performance. Problem-solving strategies, metacognition, beliefs, and domain-specific practices are also aspects of mathematical behavior. These are essential components of a theory of mathematical behavior. And it's not just mathematics; a good case can be made that they are relevant in any domain" (p. 444).

College mathematics courses for pre-health students provide them with necessary mathematical knowledge and skills that are needed for them to succeed in health-related fields of study and careers. However, current assessment procedures in this area are mostly summative or terminal in nature which provides students with a final letter grade. Developing better learning opportunities for students has not been considered seriously in the evaluation procedure. Therefore, providing constructive feedback to students about their strengths and weaknesses is hard. To overcome these difficulties, this paper proposes a self assessment component for evaluating student strengths and weaknesses for two

reasons. First, this exercise will provide learners with better learning opportunities by diagnosing their own learning difficulties. Secondly, when they enter their prospective careers in the future, they will be more comfortable with self assessment procedures which are common to health-related fields nowadays.

In terms of the structure of pre-health mathematics courses, students are expected to have three basic skills when they are solving a mathematical problem. First, students should be able to read the problem and identify an appropriate strategy to be applied in the solving process. Secondly, they need to apply the skills such as use of graphs, constructing tables or charts, use of mnemonics, or use of calculators to carry out the selected strategy. Thirdly, students should be able to explain their procedures and solutions using words thereby improving their communication skills. In other words, the problem solving process involves reading and identifying the unknowns in the problem, translating the problem into a symbolic, diagrammatic, tabulated, or any other form to facilitate the solving process, applying mathematical techniques to solve the problem, and interpret the results. In this whole process, students may need some additional information processing skills such as use of tables, graphs, charts, or mnemonics. In addition, a student has to have some technological skills such as using a calculator to perform calculations.

The above guidelines support to identify the important components as criteria which should be included in the evaluation rubric. In my opinion, an analytic rating rubric with ungrouped, multiple points will serve to assess all the above categories. Arther & McTighe (2000) express the importance of using analytic rubrics for learning improvement. According to them,

> "Analytic trait systems are not worth the effort in the classroom if all they are to be used for is putting grades on student papers. If, however, they are used as an instructional methodology - to focus instruction, communicate with students, allow for student self-evaluation, and direct instructions on traits – they are very powerful" (p.24).

The reason for choosing this kind of a rubric is that this rubric type is very informative for the teacher and it provides a detailed analysis of learner difficulties. Therefore, this type of a rubric as a diagnostic tool can be used to improve teaching and learning. Training of users to use any instrument improves its validity. Abiding to this principle, it is better to train the students beforehand for the actual self assessment exercise and also to change the criteria or other components through discussions among them if that need arises. In this way, it is expected to increase the level of learner empowerment.

Using learner-constructed criteria for self assessment has advantages as well as disadvantages. According to Orsmond et al. (2000), allowing students to construct there own marking criteria may lead to different learning outcomes compared to providing students with marking criteria. However, one disadvantage of this approach is that students may be less able to discriminate between marking criteria that they have constructed by themselves compared to marking criteria that are imposed on them. As another disadvantage, students may form integrated 'mind maps' of the marking criteria during the construction of them, and do not see the marking criteria as discrete terms as given criteria. I argue that training of users will solve this problem to some extent. However, according to Orsmond, asking students to construct their own marking criteria in discussion with their tutor and fellow students does not enhance agreement of student/teacher or student/student marking. Further, such agreement should not be seen as a measure of the success of peer and self-assessment exercises.

The main purpose of this assessment exercise is to improve student learning. It is, therefore, appropriate to discuss briefly how this rubric would help different learning styles of students. Entwistle & Tait stated four approaches which are commonly used to describe learning styles in an educational context (Cassidy, 2007). They are deep, surface, strategic, and apathetic. The deep approach learners are: intrinsically motivated, favoring active learning, and intentional to understand. It is the strategic approach which mostly reflects an alertness to assessment demands. In adopting a strategic approach, students aim to achieve optimal performance in assessments by utilizing strategies for study organization and time management. Therefore, Entwistle & Tait's ideas tell us that self assessment criteria should vary by students' learning styles as well.

The assessment format used in the current study is more open-ended and it is more closely aligned with deep and strategic approaches. Hence, it could be anticipated that students showing preferences

for deep and strategic approaches will provide more accurate estimates of assessment than surface and apathetic learners. Also, deep and strategic learners perceive themselves as more suited to open-ended assessment formats. Therefore, they will be less inclined to underestimate their marks than surface and apathetic learners. But as a whole, the instrument will serve its purpose for all types of learners.

The assessment instrument

1. Student opinion questionnaire

Cowan in Stefani (1994) supports the idea that the benefits of self-assessment are so great that we should trust students to act appropriately even when there is a risk that there could be differences between the student mark and the tutor mark. In support of this idea, Stefani reported a study where students participating in a peer and self-assessment exercise in a biochemistry practical course were given a questionnaire to evaluate the learning benefits. The results showed that almost 100% of the students said that the scheme made them think more, 85% said it made them learn more and 97% said that it was challenging. These responses were given despite the fact that 100% of the students said that it was more time consuming and over 75% said that it was hard.

Although the above results were from a study that has been designed for grading purposes, these results could comparatively be applicable to improve learning. Based on the above arguments, it is evident that one way of increasing the validity of self assessment is to make learners feel comfortable and/or believe in self assessment. It is, therefore, a good practice to include a component to self assessment which measures learner attitudes about the process. The appropriate time to administer this kind of an instrument is immediately after the initial training of students. A simple attitude questionnaire for the learners to rate their attitudes about self- and peer-assessment is proposed here. After the initial training, students will be provided with an opportunity to express their opinions by rating the following two attitude scales. Table 1 includes a rating scale for students to rate their opinion about their self assessment while table 2 provides a rating scale to rate opinions about peer assessment.

Table 1: Rating scale of student opinions about self assessment

Please rate your opinion about the self assessment exercise in the following dimensions.

Fair				Unfair
Uninformative				informative
Aids learning				Doesn't aid learning
Inaccurate				Accurate
Unbiased				Biased
Easy				Hard
Good				Bad
Makes me critical				Makes me uncritical

Table 2: Rating scale of student opinions about peer assessment

Please rate the peer assessment exercise in the following dimensions.

Peer assessment makes me:

Independent				Dependent
Not think				Think
Learn more				Not learn more
Lack confidence				Confident
Critical				Uncritical
Informative				Uninformative

2. The evaluation rubric

By examining the results of the above opinion survey and if students are satisfied with the self assessment and peer assessment process, the next step is to provide them with an evaluation rubric. Based on the essential employability skills and the essential components of a mathematical problem solving task discussed in the previous sections, the evaluation rubric in table 3 is proposed for self assessment and peer assessment of mathematical problem solving tasks. The rubric contains four main evaluation criteria each having five levels of performance from 0 to 4. A zero indicates the lowest level of performance while a four indicates the highest level. Since the main objective of this rubric is to improve student learning, a third column was added to the rubric which indicates student's weaknesses, strengths, and/or needs under each criterion level. This will help students to draw ideas from this column later to complete the assessment schedule in instrument 3.

Table 3: Rubric for assessing mathematical problem solving tasks

Scoring criteria	**Performance standards**	**Strengths and/or needs**
Conceptual understanding and connections	0 – No conceptual understanding 1 - Found almost no important math terms or their connections 2 – Showed some understanding of math terms and their connections 3 - Used most math terms correctly and showed an understanding of their connections 4 - Used math terms correctly and showed a complete understanding of how they connect	
Strategies and reasoning	0 – No plan shown 1 - Showed a plan that is not reasonable or with unnecessary information 2 - Showed some of the steps but the plan was not clear 3 - Showed a reasonable plan and most of the steps used to solve the problem 4 - Showed all the steps used to solve the problem	
Computation/Execution	0 – No computation 1 – All the computations are wrong but attempted all or some of them. 2 - Major errors in computation and arrived at a wrong answer 3 - Minor errors in computation 4 - Computed with no errors	
Communication	0 - No written explanation 1 - Communicated something that didn't go with the answer 2 - Communicated something about what was done or why it was done but not both 3 - Communicated mostly about what was done and a little about why it was done 4 – Completely communicated what was done and why it was done	

The description of each performance level in terms of student performance is : 0 – Below expectations, 1 – Marginal, 2 – Emerging, 3 – Developing, and 4 – Developed.

3. Self assessment schedule

Since the main intention of completing the rubric is to get an idea about the student's current level of performance, the next step is to plan future action. Partly, this can be done by using a self assessment schedule. The aim of a self assessment schedule is to provide the learner with a comprehensive and analytical record of learning. It is a personal self reporting on learning which can be used in the present context as a product of learning and achievement. Boud (2003) defined a self assessment schedule as a document which records student's goals and achievements in a given area and judgments about them. It is a statement which needs to contain sufficient information in it to enable someone who is familiar with the general area of the subject to ascertain what learning activities the author has engaged in and what he or she has learned. Based on Boud's above comments on a self assessment schedule, he proposes five components that have to be included in such a schedule. Goals include goals relating to the process as well as the outcomes of the course. Criteria represent the yardsticks against which it is possible to judge whether the goals were achieved. Evidence indicates what evidence learners do have for the pursuit and attainment of their goals. Judgments include a qualitative analyses of the extent to which objectives had been met and further action means future actions which have to be taken to pursue objectives.

This kind of a schedule will provide students with opportunities to reflect on their past learning experiences and at the same time think about their future actions. Considering the above arguments, it is appropriate to include a self assessment schedule for students to complete after the rubric. The contents of this schedule are given below. While the first four sections of the schedule could be started immediately after completing the self assessment exercise, the results section can be completed at a future date indicating overall improvement.

Goal – What do I want to learn?

Location – Right now I can do……..

Plan – My goal is to get ……before the next……. I need to improve in …….

Action – I'll start in…….. I'll use this strategy to……..

Results – Did I follow through with my plan? What happened? Is there any improvement?

Summary

Student involvement in their own assessment can add reflection and metacognition to the learning process. Based on this idea, the development of the stages of the assessment instrument is discussed in the concept map in figure 1. As mentioned previously, the main objective of this self assessment exercise is to improve student learning and it will not be used for grading purposes. This led to decide on which rubric type could be best suitable. An analytic, multiple-point rubric was thought to be appropriate as this type would provide detailed diagnostic features of the process. Since the rubric was designed to assess student problem solving abilities in mathematics, the domain of assessment was cognitive. Studies found that peer assessment is more accurate than self assessment. Therefore, it was decided to include a peer assessment component as well in the assessment process to gain more accuracy. Considering learner involvement in deciding evaluation criteria and other features, this rubric falls into the technical/objectified level of learner empowerment.

Many problem solving tasks involve a common set of skills such as conceptual understanding, strategy selection, reasoning, and communication. Based on the above categories, four main criteria were decided. One of the aspects that has to be looked at when constructing an evaluation instrument is the ways to increase its validity. One way of doing this is to train the evaluators before the actual exercise. Training is an essential feature for students to get familiarized with the instrument and its features. After the initial training, an attitude questionnaire was designed to be completed by students to obtain their satisfaction level about the assessment process. Further, selecting valid criteria which measure

the components of the main activity will also increase instrument validity. There is no specific point of time that this assessment could be carried out since the objective is to improve learning. Therefore, the timing of the process was considered to be formative and ongoing. Ultimately, there should have to be some evidence and a future plan for learners to improve their areas of weaknesses. An assessment schedule was considered to be the best kind of solution for this.

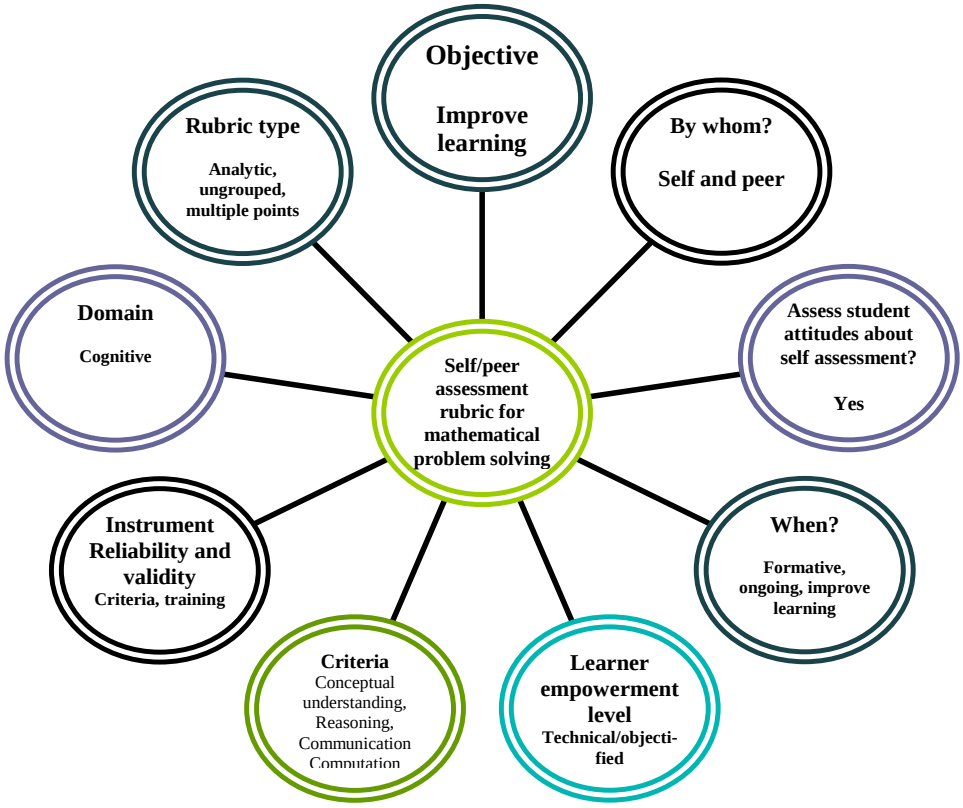

Figure 1: A concept map indicating the initial thinking process involved in the research paper

Discussion

In many classroom situations, rubrics are used to assess and grade students. Rubrics are not only scoring tools, but also important instructional tools. Many students are unaware that rubrics help them to improve their learning. Thus, it could be challenging at the beginning to get students involve in the process as many of them always value a letter grade as an end product. However, when proper training is given and once the students know the value of self assessing for the improvement of learning, they will be more involved in the process and intrinsically motivated than ever before. Developing good usable rubrics is in itself challenging. However, well-developed rubrics are not enough. Users still need to be given training and guidance on the use of rubrics. Such training and guidance will not totally eliminate all discrepancies, since any judgment based on rubrics will not be absolute. But this will help reduce the number and size of discrepancies.

In a course with external limitations on grading, there are always problems with negotiating an assessment procedure. When the assessment policy is unilaterally decided by the institution policy, it is quite challenging to obtain the real benefits of a self assessment exercise. Also, most students find current norm-referenced grading schemes to be educationally unsound. Further, an assessment procedure takes a great deal of class time, partly because of the anxiety many students have about the assessment. Many of the above issues do not arise in the current exercise as the assessment procedure

has purely designed for the improvement of learning. Therefore, one of the effective ways of improving learning is to enhance students to self/peer-assess their strengths and/or weaknesses. The guidance provided in this paper explains how this process could be carried out safely and securely. In my opinion, teachers should be more encouraged to do this type of work for their students. However, it is important that students should be aware of the objectives of the exercise and a prior training for them is not only important but essential as well.

On the other hand, since there is no assessment part for the teacher in the scheme suggested in this paper, student anxieties about teacher-grading will diminish thereby providing better learning opportunities for students through the exercise. The approach developed in this paper is one, which could be satisfactorily used in many situations without any preferred terminal grades. By this way, students will have a significant degree of responsibility on their own learning. Also, this exercise has a potential for encouraging reflection, integration of learning, and most importantly metacognition. Unless students have made a commitment to learning, as distinct to complete a particular course, they will turn anything into an exercise which is graded. This form of obtaining a final grade has not been considered to be an important aspect of the proposed self/peer assessment exercise. The four stages of the self assessment process, that is self-reporting, self-analysis, self-valuation, and self-determination have been considered here to be the most important aspects of better learning.

References

Arther, J. & McTighe, J. (2000). *Scoring rubrics in the classroom: using performance criteria for assessing and improving student performance*, Corwin Press, Inc.

Arthur, H. (1995). Student self-evaluations: How useful? How valid?, *International Journal of Nursing Studies, 32(3)*, 271-276.

Birenbaum, M., & Dochy, F. (1996). *Alternatives in assessment of achievement, learning process, and prior knowledge*. Boston, MA: Kluwer.

Boud, D. (2003). *Enhancing learning through self assessment*, RoutledgeFalmer.

Cassidy, S. (2007). Assessing 'inexperienced' students' ability to self-assess: exploring links with learning style and academic personal control, *Assessment & Evaluation in Higher Education, 32(3)*, 313-330.

Conference Board of Canada (1992). *Employability Skills Profile (Draft)*, Lecture notes, CTL 1799 – Self assessment, winter 2008.

Hendry, G. D. (1996). Constructivism and educational practice. *Australian Journal of Education, 40*(1), 19-45.

Lam, T. (2008). *Lecture notes, CTL 1799 - Self assessment*, winter 2008.

Lane, S. (1993). The conceptual framework for the development of a mathematics performance assessment instrument, *Educational Measurement: Issues & Practices, 12(2)*, 16-23.

Manning, S. (1993). Intelligence: Let us count the ways, *Research speaks to teachers*, 28(2).

Meier, S. L., Rich, B. S. & Cady, J. (2006). Teachers' use of rubrics to score non-traditional tasks: Factors related to discrepancies in scoring, *Assessment in Education, 13(1)*, 69-95.

Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment, *Assessment & Evaluation in Higher Education, 28(4)*, 383-394.

NCTM (1989). *Curriculum and evaluation standards for school mathematics*.

Nitko, A. J. (1996). *Educational assessment of students*, Second edition, Prentice-Hall Inc.

Orsmond, P., Merry, S. & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment, *Assessment & Evaluation in Higher Education, 25(1)*, 23-38.

Popham, W. J. (1997). What's wrong - and what's right – with rubrics, *Educational Leadership, 55(2)*, 72-75.

Quinlan, A. M. (2006). *A complete guide to rubrics: Assessment made easy for teachers, K-College*, Rowman & Littlefield Education.

Sambell, K. & McDowell, L. (1998). The value of self and peer assessment to the developing lifelong learner. In C. Rust (Ed.), *Improving student learning – improving students as learners* (pp. 56-66). UK: Oxford Centre for Staff and Learning Development.

Schoenfeld, A. H. (2002). *Research methods in mathematics education*. In L. D. English (Ed.), Handbook of International Research in Mathematics Education, Lawrence Erlbaum Associates (pp. 435-487).

Sluijsmans' D. M. A., Moerkerke, G, Merrienboer, J. J. G.V. & Dochy, F.J.R.C. (2001) peer assessment in problem based learning, *Studies in Educational Evaluation, 27*, 153-173.

Stallings, V. & Tascione, C. (1996). Student self-assessment and self-evaluation, *The Mathematics Teacher, 89(7)*, 548-554.

Stefani, L. A. J. (1994). Peer, self, and tutor assessment: relative reliabilities, *Studies in Higher Education, 19(1*), 69-75.

Watson, A. (2006). Some difficulties in informal assessment in mathematics, *Assessment in Education, 13(3)*, 289-303.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*, Jossey-Bass Publishers.

Author

Gunawardena Egodawatte, Department of Curriculum, Teaching and Learning, University Toronto, Toronto, ON, Canada; e-mail: egunawardena@oise.utoronto.ca