ASSESSMENT IN THE MATHEMATICS CLASSROOM

Yearbook 2011

Association of Mathematics Educators

This page intentionally left blank

ASSESSMENT IN THE MATHEMATICS CLASSROOM

Yearbook 2011

Association of Mathematics Educators

editors

Berinderjeet Kaur

Wong Khoon Yoong

National Institute of Education, Singapore

World Scientific

AME
ASSOCIATION
OF MATHEMATICS
EDUCATORS

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ASSESSMENT IN THE MATHEMATICS CLASSROOM
Yearbook 2011, Association of Mathematics Educators

ISBN-13 978-981-4360-97-5
ISBN-10 981-4360-97-X

Printed in Singapore.

Contents

Chapter 1

Introduction: Assessment Matters

WONG Khoon Yoong   Berinderjeet KAUR

This introductory chapter provides an overview of the subsequent chapters and relates them to significant issues about assessment in mathematics education relevant to Singapore and other countries. It ends with a set of reflective questions that teachers might think about while reading the chapters and thinking about making changes to their practices and beliefs about assessment.

1 Why a Yearbook on Assessment?

Ordinary yearbooks, also called annuals, normally report significant events and statistics about the past year, but academic or scholarly yearbooks in education tend to be issue-based. These issues may or may not reflect discussion that has taken over the past year by members of that scholarly community, yet the issue is important enough for the community to produce a timely collection of articles to address the issue in depth. One of the longest running education yearbooks is the Yearbook of the National Society for the Study of Education (NSSE)[1], which began publishing in 1901 and is now published under Teachers College, Columbia University. Its 2007 yearbook is entitled "Evidence and Decision Making" and it focuses on the roles that education professionals, including researchers, teachers, and administrators, can play "in constructing, interpreting, and using evidence to make decisions that support learning" (Moss and Piety, 2007, p.2), referring also to the

[1] http://nsse-chicago.org/Yearbooks.asp

"data-driven" movement. In mathematics education, the National Council of Teachers of Mathematics (NCTM) has published yearbooks since 1926, and the purpose is to "focus concerted attention on timely issues by viewing them in depth, from multiple perspectives, through interdisciplinary lenses, and across different grade bands"[2]. Its latest, the 72nd yearbook published in 2010, is about mathematics curriculum, but its chapter 6, by Peter Kloosterman and Crystal Walcott, is an example of using national assessment in the United States to investigate the link between what we teach and what the students learn. The NCTM's 1993 yearbook was on assessment in the mathematics classroom, and this might have laid the groundwork for the assessment standards announced two years later (National Council of Teachers of Mathematics, 1995).

The Association of Mathematics Educators has followed this trend of producing yearbooks based on major issues of interest to the Singapore teachers and educators. The first two yearbooks were about Mathematical Problem Solving (2009) and Mathematical Applications and Modelling (2010). This yearbook, the third one in the series, focuses on assessment. This theme follows naturally from the earlier two volumes because, having implemented problem solving, applications, and modelling in the classrooms, the teachers should then focus on assessing how well these attempts have been successful in improving mathematics learning for the students. Furthermore, there is a need to encourage teachers to include assessment of non-cognitive attributes and to use techniques in addition to standard paper-and-pencil tests that focus on typical problems. In fact, the first Mathematics Teachers' Conference organised by the Association in 2005 was on assessment, and this yearbook can be considered an update on the same theme. In his keynote address delivered at that conference, Professor David Clarke discussed how international educators could implement assessment in many different ways to serve common visions. Readers will be pleased to read his chapter (7) to find out his latest view about the advantages and disadvantages of using open-ended tasks in mathematics assessment, six years after his earlier address.

---

[2] http://www.nctm.org/catalog/productsview.aspx?id=98

The authors of the 12 peer-reviewed chapters have been asked to focus on evidence-based practices that school teachers can experiment in their lessons to bring about meaningful learning outcomes. These chapters are organised into two major sections covering assessment in the cognitive and affective domains. Most of the authors explain practical ideas that the teachers can implement in their classrooms, but they also point to existing research of their own or in the literature to support these practices. Many examples are instructive, some are promising, while others hold potentials that need to be further developed and investigated. Our task here is to provide a roadmap to the chapters by briefly summarising the main points of the chapters and using them to raise related assessment issues not covered by the authors. This will widen the scope of coverage and analysis, hopefully to stimulate the readers to ponder and to form discussion groups to explore important issues based on the authors' work. Of course, the readers can choose to start with any chapter they are particularly interested in.

A recurring theme in most chapters is the widely circulated notions of formative assessment and assessment *for* learning. Michael Scriven (1967) was credited as the first educator to distinguish between formative *evaluation* and summative *evaluation* (some writers use *assessment* and *evaluation* interchangeably, while others distinguish between them), when he was focusing on evaluation of curriculum materials. The theoretical conception of formative assessment is further developed by Sadler (1989, 1998) and others. The evidence for formative assessment becomes well established through the comprehensive and widely cited review by Black and Wiliam (1998), and their *Inside the Black Box* series. Wiliam (2007) provides an updated review specially for mathematics assessment. According to Popham's levels of implementing formative assessment (2009), teachers who make instructional adjustments after examining assessment-elicited evidence are at level 1, and teachers who train students to adjust their learning tactic after obtaining assessment results are at level 2. These and similar writings have led to the approach called assessment *for* learning (shortened to *AfL* or *A4L* in some writings). An internal review report about assessment by the Singapore Ministry of Education (MOE) in 2009 calls for a balance between assessment *of* learning with assessment *for* learning and points out the need to raise assessment literacy of its teachers and education officers. In the following sections, we will indicate how the chapters may contribute toward this important assessment goal.

2 Assessment of Mathematics Cognitive Domain

The mathematics cognitive domain, as for all other school subjects, is multi-dimensional. *Mathematics understanding* is a term widely used in a vague and generic sense to refer to what students know about a piece of mathematics, including the skill or procedure component. When students make some mistakes in carrying out a procedure, teachers often claim that they have not *understood* the mathematics, failing to explicate what this means as well as confounding the distinctions between concepts and skills. This vagueness poses some problems for valid assessment of *mathematical understanding*. This has been a controversial issue at least since the publication of the original taxonomy of cognitive domain by Bloom (1956), but in mathematics education, the nature of *mathematical understanding* has been subject to extensive analysis, for example, it has been classified as instrumental and relational understanding, know what, know how, and know why, and other possibilities (e.g., Schoenfeld, 2008; Sierpinska, 1994; Skemp, 1979; Wong, 1985). In chapter 2, Thompson and Kaur propose the SPUR approach to unpack the multi-dimensions of *understanding* by focusing on **S**kills, **P**roperties, **U**ses, and **R**epresentations. They provide examples of SPUR at the primary and secondary level, and teachers may be able to design similar tasks "to help students develop a well-rounded and balanced view of a given topic" (p. 25). The authors then report the results of a comparative study called the International Project on Mathematical Attainment (IPMA) of a sample of Singapore and US primary school pupils in these four dimensions by topics (Number, Algebra, Measurement, Geometry, Data/Chance). The relatively few items in IPMA covering Properties and Uses suggest that it is harder to design such items compared to Skills and Representations. This highlights an important area for teacher professional development. Thus, their chapter contributes to the ongoing discourse about the nature of *understanding*, an ubiquitous and all-embracing construct.

Problem solving has been the central focus of the Singapore mathematics curriculum framework since the early 1990s, and since then more than 60 local studies have been conducted about this important area, some of which were reviewed by Foong (2009). What will be of particular interest to teachers are new ways teach and assess mathematical problem solving. Toh, Quek, Leong, Dindyal, and Tay, in chapter 3, introduce their so-called “mathematical practical worksheet” as both a learning and an assessment instrument. This worksheet covers the four stages based on the influential work of Polya and Schoenfeld: understand the problem; devise a plan; carry out the plan; check and expand. A rubric is provided to score student work at each of these stages by levels, and students are made aware of these assessment criteria. Two student examples are discussed, one a high-achieving student while the other a “non-believer” in problem solving. Extensive transcripts of the interviews with these two students are used to interpret the stages of problem solving. They conclude from their research with 24 Year 8 students that this worksheet approach is a promising one, stressing the need to make the problem solving process and the assessment criteria transparent to the students. Indeed, knowledge of assessment criteria is one way to include students in the assessment process called self-assessment, but to strengthen this approach, teachers and students should discuss their interpretations of the criteria because it cannot be assumed that both parties will interpret the criteria in the same way.

Conceptual understanding is a critically important outcome, without which students are likely to resort to memorising procedures in a mechanical and meaningless fashion. To advance in mathematics learning, students need to know the inter-relationships of key concepts, for example, an *equation* is built from *expressions,* and it leads to types such as linear equations, quadratic equations and so on, satisfying the central concept that its solutions are values that make the equation into a true statement. However, typical paper-and-pencil mathematics tests designed by school teachers and used in public examinations in Singapore concentrate on solving routine and non-familiar problems. These mathematics tests rarely include assessment of conceptual links. In chapter 4, Jin and Wong argue that it is important to assess conceptual understanding, and they describe how this can be achieved using concept map, which has strong research support and extensive practical exemplars in science education. They describe in detail a training programme to help teachers prepare their students to use concept mapping in learning and three types of concept mapping tasks that can be used for assessment: fill-in-the-map, semi-directed concept mapping, and free-style mapping. Several examples from elementary algebra and geometry are given, and simple methods to assess the links across concepts are discussed. Since it is time-consuming to learn and assess concept maps, teachers who intend to use this technique will be better served if they think about using concept mapping as both assessment *of* learning (with scoring as illustrated in the chapter) and assessment *as* learning (students strengthen their conceptual understanding as they think deeper about the concepts when they try to produce the concept maps).

Besides paper-and-pencil tests, teachers should acquire other types of evidence about students' mastery of mathematical contents, processes, and applications. Two chapters describe several so-called *alternative* assessment methods, which could be better called *additional* assessment modes because they are unlikely to replace the traditional problem solving tests.

First, Kaur and Chan in chapter 5 describe student journal writing as a complementary source of evidence. They provide a set of prompts for cognitive as well as affective outcomes. Examples of student work cover shapes, area and perimeter, and factors and multiples. An analytic scoring rubric and a holistic scoring rubric are described for the cognitive items, whereas journal writing about affect does not need to be graded. The authors briefly caution teachers about four potential pitfalls if this technique is used mindlessly: hurt students' feelings, loss of instructional time, increase in marking workload for the teachers, and deciding to grade mathematics contents or language. In fact, the advantages of such journal writing can be strengthened if timely and specific feedback can be given to the students. To enhance their own assessment literacy, teachers might wish to form discussion groups to design a feedback system for the samples included in this chapter. A pragmatic feedback technique is to give only written comments. Indeed, research has shown that giving only written comments may lead to higher achievement than giving grades or giving grades with comments, this last finding is really surprising and counter-intuitive (see Wiliam, 2007).

Second, Yeo in chapter 6 describes four types of additional assessment methods for lower primary classes: practical tests, oral presentation, journal writing, and open-ended tasks. Instructions are given, but teachers need to experiment with these tasks in their classes and reflect on how well they fit into the recently implemented "holistic assessment" for Singapore primary schools, recommended by the Primary Education Review and Implementation (PERI) Committee and first piloted with 16 PERI Holistic Assessment Prototype (or HAPpe) schools from July 2009[3]. A range of "bite-sized" forms of assessment are to be used to replace end-of-year examinations after 2013, and these class tests or quizzes might include the four types enunciated in this chapter. Indeed, the teachers have to know how to use the different evidences gathered from these assessment modes in order to plan engaging lessons to help their pupils learn better.

We wish to comment briefly on the use of rubrics to assess mathematical knowledge. In 2004, the Ministry of Education launched the SAIL (Strategies for Active and Independent Learning) project. It is an instructional approach in which students are made aware of the expectations of learning a topic in terms of how learning tasks are to be assessed based on a rubric. For mathematics, the assessment rubric covers the first four criteria for primary schools and all five for lower secondary level: Approach and Reasoning; Solution; Overall Presentation; Connections; Mathematical Language and/or Representation. Each criterion is to be assessed from Level 1 to Level 4, and the resulting assessment matrix of Criteria × Levels will provide a comprehensive map of student learning outcome. This assessment tool is supposed to promote "a learning environment that is characterised by learner-centred processes, dynamic classroom talk, and differentiated learning" (Ministry of Education, Singapore, 2004, p.15). This SAIL framework and the rubrics designed by the above authors should provide

---

[3] http://www.moe.gov.sg/media/speeches/2010/07/13/peri-holistic-assessment-seminar-2010.php

good examples for teachers who plan to use a similar technique in their teaching.

The earlier chapters have explicitly or implicitly touched on the openness of some mathematics assessment tasks, and it is necessary to subject this construct to serious analysis. Clarke, in chapter 7, has most thoroughly accomplished this by discussing the positive and negative reasons about the use of open-ended tasks to assess student mathematics learning. He urges teachers, schools, and school systems to systematically consider these "answers" when they develop assessment schemes for their students. He highlights three challenges: identify criteria used to select open-ended tasks (he has given several examples), train and support teachers and students in using these tasks, and find constructive ways to make good use of the data. In order to tackle these challenges, all the participants in this complex assessment endeavour must boldly grasp the nettle, as his title suggests, for it is only when steps are taken boldly, firmly, and immediately that there are good chances of success.

Over the past three decades, advances in computer technologies and software coupled with large-scale assessments and accountability requirements in some countries have led to the development of technology-based tools to collect, manage, and use assessment data to inform instruction (Haertel, Means and Penuel, 2007). In chapter 8, van den Heuvel-Panhuizen, Kolovou, and Peltenburg describe three ways to use ICT to promote better assessment of student learning: (a) use tasks of high mathematical demand and complexity, (b) make tasks more accessible to students so that there is better opportunity for them to demonstrate what they know, and (c) track students' thinking and solution processes, using the novel approach of screen video and log files of student activities. Their examples are truly inspiring, but it requires a strong team of members with different expertise to create similar examples and analysis tools. Hopefully in the not too distant future, similar projects can be carried out in Singapore, with collaboration among the Educational Technology Division of the MOE, NIE, and the schools.

For the past two years, the MOE has articulated the C2015 Student Outcome framework, under which students are expected to become self-directed learners, confident persons, active contributors, and concerned citizens. Teachers are tasked to provide the environments and tools, in particular ICT ones, to promote self-directed learning among their students. Thus, there is also a need to develop assessment schemes to evaluate the success or otherwise of this kind of learning. We are indeed fortunate to have in this book a substantial chapter 9 contributed by Mok on her comprehensive project called Self-directed Learning Oriented Assessment (SLOA), which has been successfully implemented in about 130 schools in Hong Kong, Macau, and China. This extensive project covers the three aspects of assessment: *of* learning, *for* learning, and *as* learning, and the potentials for self-directed learning can be induced in all three aspects, in particular the *as* learning notion, in the form of a learning goal. The main bulk of the chapter is on assessment *for* learning. Her project involves creating tens of thousands of items that match the national mathematics curriculum in Hong Kong and analysing them with sophisticated techniques. Two analyses are explained: item-person maps obtained from the Rasch model that can be used to identify a student's zone of proximal development, and the student-problem charts used to identify anomalies in responses among a group of students. Teachers first need to learn how to interpret student results reported using these two techniques and then how to generate similar results from their own tests and students. Needless to say, ordinary teachers are unlikely to be able to carry out such an intricate assessment mode on their own, so a whole school approach that includes external experts is required, as Mok has stressed the SLOA school culture in her chapter. Readers who wish to learn more about SLOA should consult Mok's latest book (2010).

Public examinations have been the stable feature of the Singapore education system, since all students have to take them at the end of their primary and secondary education. However, not all Japanese students have to take national examinations in Mathematics as in Singapore, and in the past years only samples of Japanese students at different grades take large-scale examinations. Chapter 10 by Shimizu explains this Japanese experience. He focuses on the framework and sample items taken from the *National Assessment of Academic Ability and Learning Environments,* conducted for the first time in April 2007 for only grades

6 and 9 students. Of particular interest are the use of real-world contexts (students are required to think about numbers and figures when they observe things in their surroundings) and the ability to explain their observations, solution methods, and reasoning (including proofs) in the open-construction tasks. To help teachers become more familiar with these new tasks, the Japanese Ministry of Education, Culture, Sports, Science and Technology has circulated leaflets about the items to the schools. It also provides sample lesson plans to show teachers how to use data from this large-scale assessment to move students' thinking forward, in an effort to bridge the gap between large-scale external assessment and the classroom practices. This approach to use student results as assessment *for* learning carried out at the national level, albeit in a "delayed" rather than immediate way, is a model worthy of emulation. This can be taken as an example of what Clarke refers to as a "social contract" in which the assessing body (usually the assessment section of the ministry) provides teachers with support and guidance to use quality assessment data to improve teaching. In Singapore, there must be huge amount of data about student mathematics performance in public examinations over the past several decades, but we do not know of any attempt to share the data with school teachers and to suggest classroom lessons that can improve student learning. This social contract needs to be looked into.

The quality of assessment of mathematics contents depends on the validity of the items, irrespective whether these are designed by the teachers or examination boards. It is at the pre-service training stage that future teachers learn to construct valid items. Pre-service teachers must be strong in mathematical content knowledge so that they can construct items that are error-free. Dindyal in chapter 11 compiles a set of errors made by pre-service teachers at the NIE and relates those errors to language, mathematics content, diagrams, and contexts (or problem situations). These errors are handy signposts to remind teachers to be more careful when they design similar items for their own use, because poorly constructed items can lead to wrong assessment of what students actually know and can do. Indeed, the ability to design traditional construct-response or select-response mathematics items is a skill to be developed over a considerable period of time, and one effective way is for teachers to critique items created by colleagues within a supportive working environment.

3 Assessment of Mathematics Affective Domain

There are only two chapters about assessing students' mathematics affective attributes or outcomes. However, this should not be taken as indicating that student attitudes, motivations, efficacy, and similar attributes are not important. Indeed, the Singapore mathematics curriculum framework has included *Attitudes* as one of the five key factors to help students become successful problem solvers. The framework specifically lists these five attributes as desirable: beliefs, interest, appreciation, confidence, and perseverance. It calls on mathematics teachers to design learning activities that are fun, meaningful, relevant, and likely to help build confidence in the students. The two chapters here will provide different ways that teachers can use to measure these student attributes.

This section begins with a succinct summary (or a "quick start") of affective assessment by Tay, Quek, and Toh in chapter 12. After arguing for the inclusion of affective assessment in the mathematics classrooms, they describe with examples three modes of assessment: summated scale, interest inventory, and semantic differential. Although they agree with Popham (2009) that valid affective assessment should be about the whole class rather than individual students, part of the reasons being that current affective assessment tools are not sufficiently accurate to measure attributes of individual students, this is not what most teachers and researchers do. Teachers wish to know something about individual students so that they can provide encouragement, feedback, and follow-up activities targeted to the individuals. On the other hand, researchers are interested to study the relationships between cognition and affect, and this kind of correlation studies require matching data collected about individual students. This is an issue that teachers need to ponder over, especially after they have gained some experience in using the techniques described by these authors.

To help teachers include self-assessment into the learning of mathematics, Fan begins chapter 13 by linking self-assessment to metacognition in the Singapore mathematics curriculum framework and then provides examples of worksheets to cover structured self-assessment, integrated self-assessment, and instructional self-assessment. He refers the readers to his research about self-assessment conducted under the Mathematics Assessment Project (MAP) around 2006.

Other than these two chapters, earlier chapters on student journal writing and reflection sheets with prompts can be readily modified to capture evidence about affective learning in addition to cognitive outcomes. The examples are currently based on hand-written responses. However, as more and more students grow up as digital natives, it is necessary for teachers to experiment with Web 2.0 tools such as blogs and RSS (Really Simple Syndication) readers as medium to deliver, accept, manage, and evaluate student journal entries and responses to checklists (Marino and Angotti, 2011). Teachers who have gained experience in this area are welcome to contribute papers for *The Mathematics Educator* and the *Maths Buzz*.

4 No "Final" Words: A list of Questions

Years of policy decisions, school practices, and research about assessment have generated a large corpus of knowledge all over world that teachers can tap into to guide their practices. However, many questions remain unresolved, not because of lack of attention; but rather it is the complexity of human learning and teaching and the ever increasing changes in society that have had unpredictable impacts on school practices. Below are some questions about assessment that we think teachers may consider when they develop their assessment literacy, make decisions about changes to their assessment practices, and conduct classroom research as part of their professional development. Some tentative "answers" may be found in the subsequent chapters!

1. When teachers are advised to adopt a new assessment mode and technique, they should ask themselves some searching questions. Why do we want to adopt this technique? Is it to replace or to supplement current assessment practices? Does it fit into a coherent assessment

scheme that aligns with school priorities and the national curriculum? How much time and effort is required to master the technique so that it can be used as part of daily instruction rather than infrequently, say once a term?

2. If a new technique is to be implemented as assessment *of* learning, will the grading be valid, reliable, fair, and easy to use? What additional dimensions of cognition or affect can it assess? How can different assessment results be used to map student's progress through the grade levels and in specific mathematics contents?

3. If a new technique is to be implemented as assessment *for* learning, does it enhance or hinder students' opportunities to learn mathematics? What follow-up activities are appropriate for different types of responses?

4. "Teaching to the test" or "what are tested will be taught" is a common phenomenon all over the world including Singapore. This is often seen in the practices where teachers let students practise typical questions that appear in high-stake public examinations. Are there strong theoretical reasons and empirical data to support or to reject this approach? To what extent do you subscribe to this claim in beliefs and practices?

5. What kind of professional training and support do teachers need to be able to implement the new technique with confidence?

6. What additional demands are placed on the students? What do they gain by being assessed using a new technique? What will they lose?

7. Will the inclusion of additional assessment data be used to evaluate teacher's instructional quality and effectiveness?

Sufficient details must be given to guide teachers toward how to implement a new assessment technique; see examples in Ellis and Denton (2010). Implementing even one or two techniques discussed in the following chapters can be quite overwhelming for the busy teachers. The challenge at the systemic level is for the assessment agency to plan several coherent assessment schemes that teachers can adapt, and study their efficacy. While it is understandable that advocates of new techniques are keen to highlight the benefits of innovations, educators should strive to present a "balanced" summary of what works and what does not work, and under what conditions. Currently innovative or alternative assessments have weak research support in Singapore because many of them are implemented infrequently for various reasons. These techniques should be implemented frequently enough to have good chances for them to work in concert to realise new curriculum goals. More well-designed studies are needed.

We leave it to the readers to use the chapters as starting points to reflect on their own practices and to engage colleagues and external experts in conversations about how they might improve their assessment literacy and practices within a professional learning community (PLC), which is strongly advocated by the MOE. We hope that this book will play a significant role in informing teachers so that their practices can be enhanced for the benefits of their students.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-75.

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: Longmans, Green.

Ellis, A. K., & Denton, D. W. (2010). *Teaching, learning, and assessment together: Reflective assessments for middle and high school mathematics and science*. Larchmont, NY: Eye on Education.

Foong, P. Y. (2009). Review of research on mathematical problem solving in Singapore. In K. Y. Wong, P. Y. Lee, B. Kaur, P. Y. Foong & S. F. Ng (Eds.), *Mathematics education: The Singapore journey* (pp. 263- 300). Singapore: World Scientific.

Haertel, G. D., Means, B., & Penuel, W. (2007). Technology tools for collecting, managing, and using assessment data to inform instruction and improve achievement. *Yearbook of the National Society for the Study of Education, 106*(2), 103-132. doi: 10.1111/j.1744-7984.2007.00117.x

Marino, K. J., & Angotti, R. L. (2011). Mathematics journals made manageable: Using blogs and RSS feeds. *Mathematics Teachers, 104*(6), 466-469.

Ministry of Education, Singapore (2004). *SAIL: Strategies for Active and Independent Learning*. Singapore: Author.

Mok, M. M. C. (2010). *Self-directed Learning Oriented Assessment*. Hong Kong: PACE Publications.

Moss, P. A., & Piety, P. J. (2007). Introduction: Evidence and decision making. *Yearbook of the National Society for the Study of Education*, *106*(1), 1-14. doi: 10.1111/j.1744-7984.2007.00095.x

National Council of Teachers of Mathematics (1995). *Assessment standards for school mathematics.* Reston, VA: Author.

Popham, W. J. (2009). *Unlearned lessons: Six stumbling blocks to our schools' success.* Cambridge, MA: Harvard Education Press.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119-144.

Sadler, R. (1998). Formative assessment: Revising the territory. *Assessment in Education, 5*(1), 77-84.

Schoenfeld, A. H. (2008). Mathematics for understanding. In L. Darling-Hammond (Ed.), *Powerful learning: What we know about teaching for understanding* (pp.113-150). San Francisco, CA: Jossey-Bass.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp.39-83). Chicago, IL: Rand McNally.

Sierpinska, A. (1994). *Understanding in mathematics*. London: Falmer Press.

Skemp, R. (1979). Goals of learning and qualities of understanding. *Mathematics Teaching, 88*, 44-49.

Wiliam, D. (2007). Keeping learning on track: Formative assessment and the regulation of learning. In F. K. Lester, Jr (Ed.), *Second handbook of research on mathematics teaching and learning* (pp.1053-1098). Greenwich, CT: Information Age Publishing.

Wong, K. Y. (1985). *Mathematical understanding: An exploration of theory and practice.* Unpublished PhD thesis. University of Queensland.

Chapter 2

Using a Multi-Dimensional Approach to Understanding to Assess Students' Mathematical Knowledge

Denisse R. THOMPSON Berinderjeet KAUR

Many educators around the world have called for curriculum and teaching that aim for a balanced perspective on procedural fluency as well as conceptual understanding. Assessment also needs to emphasize this balance. In this chapter, we advocate for a multi-dimensional approach to assessing students' understanding of mathematics, specifically their ability with skills, mathematical properties, uses or applications of mathematics, and representations of the concepts. We argue that each dimension provides different insight into students' understanding. We also share assessment results from the United States and Singapore, illustrating that overall scores often mask differences in achievement that are essential for teachers to understand if they are going to design instructional activities that will help students develop their maximum mathematical potential.

1 Introduction

Before reading this chapter further, reflect on the following:

You are trying to assess your students' understanding of decimals. Which of the following questions would you consider appropriate for your students? Why? What do you learn from one question that you might not learn from another?

- Simplify: 3.28×0.5.
- If 3.28 is multiplied by a number between 0 and 1, is the product greater than, less than, or equal to 3.28? How do you know?
- If fruit costs $3.28 per kilogram, how much should you pay for 0.5 kilogram?
- Where should you place 3.28 on the following number line?

We suggest that each of the previous questions is an appropriate assessment item for teachers to use when students are studying decimals. Each provides different insight into what students know about the concept. Together, they provide a more robust view of students' depth of understanding than would be obtained from an individual item. For instance, what should teachers make of students who can complete the computation in the first bulleted item but cannot complete the same computation in the contextual setting of the third item? Likewise, students who are not able to answer the second bulleted item are often hindered in their ability to assess whether an answer to a computation, such as that in the first bullet, makes sense. By looking at students' achievement across the four items, teachers get a glimpse into potential misconceptions that students have which may influence their ability to delve deeply into mathematics.

The multi-dimensional view of understanding illustrated by the previous four items is what we call the SPUR approach, for skills, properties, uses, and representations. Each of the previous questions fits one of these dimensions. In the remainder of this chapter, we provide a brief theoretical and philosophical perspective for the SPUR approach, share more examples of items using this approach at both the primary and secondary levels, and illustrate some assessment data from the United States and Singapore that suggest a need to analyze assessment data according to these dimensions to understand students' depth of knowledge.

2 Why Consider a Multi-Dimensional Approach to Understanding?

Many mathematics educators have recognized the importance of using multiple perspectives to assess the learning of mathematics content. For instance, Freudenthal (1983) considered the different ways in which a topic might be used and how those different perspectives lead to different understandings. In a synthesis of research about children's understanding of mathematics in the United States, Kilpatrick, Swafford, and Findell (National Research Council, 2001) defined mathematical proficiency as a tree consisting of five intertwined strands: procedural fluency, adaptive reasoning, conceptual understanding, productive disposition, and strategic competence. These strands are interconnected and interdependent, with students needing to develop competence in all five strands concurrently to have a robust understanding of mathematics. Similarly, Krutetskii (1976) showed that, at least among gifted students of mathematics, some students regularly use algebraic or analytic approaches to solve problems, while others use geometric or spatial approaches.

The views espoused by these educators are also reflected in curriculum recommendations in various countries. For instance, in the United States, the *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics [NCTM], 2000), which has guided the development of curriculum materials and state curriculum frameworks, outlines a vision for mathematics for students in grades preK-12. In particular, the standards documents emphasize the importance of a balanced perspective relative to procedural fluency and conceptual understanding. Likewise, curriculum recommendations in Singapore too have emphasized the development of mathematical skills, concepts, and processes (see Figure 1) essential in the learning and application of mathematics (Ministry of Education, 2006a and 2006b).

Given these recommendations, curriculum materials that use a multi-dimensional perspective present a balanced view of mathematics that accommodates classrooms with a range of students having different mathematical strengths and learning styles. If teaching materials reflect a multi-dimensional perspective, then assessment needs to reflect this perspective as well in order for teaching and assessment to align.

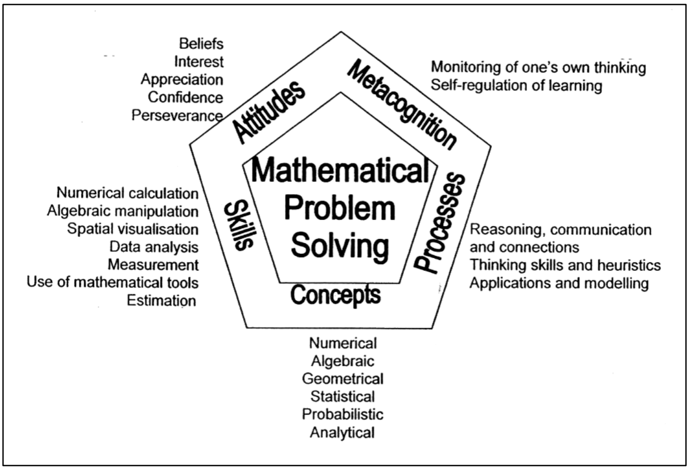


Figure 1. Framework of the school mathematics curriculum

3 What is the SPUR Approach?

One approach to a multi-dimensional perspective on understanding is known by the acronym SPUR for **S**kills, **P**roperties, **U**ses, and **R**epresentations (Thompson and Senk, 2008; Usiskin, 2003 and 2007). In particular,

> ***Skills*** represent those procedures that students should master with fluency; they range from applications of standard algorithms to the selection and comparison of algorithms to the discovery or invention of algorithms, including procedures with technology. ***Properties*** are the principles underlying the mathematics, ranging from the naming of properties used to justify conclusions to derivations and proofs. ***Uses*** are the applications of the concepts to the real world or to other concepts in mathematics and range from routine "word problems" to the development and use of mathematical models. ***Representations*** are graphs, pictures, and other visual depictions of the concepts, including standard representations of concepts and relations to

> the discovery of new ways to represent concepts. (Thompson and Senk, 2008, p.2)

We believe that students who have a robust understanding of mathematics should possess understanding in each of the dimensions, **S**kills, **P**roperties, **U**ses and **R**epresentations.

Although originally used in developing curriculum materials in the United States, SPUR can be a powerful tool for assessment as well. If assessments consistently measure students' achievement in only one dimension, then teachers may have a misconceived view of their students' understanding. In contrast, if assessments measure understanding in all four dimensions, teachers can gain insights into strengths and weaknesses in their students' knowledge of the concept that can be used to guide further instructional planning.

3.1 *Examples of SPUR at the primary level*

Consider the topic of Number, specifically fractions, which is commonly studied in the upper primary grades in many countries. Certainly teachers want students to develop proficiency with computations involving fractions. But if students are only able to compute and are not able to discuss properties related to fraction computation or use fractions in real contexts or provide visual models of fraction operations, then we would argue that their understanding is limited. Students who can view a topic from different perspectives typically have a wider repertoire of strategies at their disposal from which to draw when solving a problem.

Here is a typical fraction division task that teachers might expect students to solve with fluency:

- Simplify: $1\frac{2}{3} \div \frac{1}{6}$

But just because students can simplify the given problem using an algorithm does not mean that they have a robust understanding of division with fractions. To assess that more robust understanding, additional information is needed.

What are some of the underlying principles you would want students to know about fraction division? These principles form the basis of items focusing on the dimension of Properties. For instance, the following principles might be ones that teachers would want students to understand and/or recognize, even though primary students might not be expected to make such statements:

(1) dividing by a number between 0 and 1 results in a quotient that is larger than the dividend;

(2) the smaller the divisor for a given dividend, the larger the quotient;

(3) the larger the divisor for a given dividend, the smaller the quotient;

(4) division by 0 is undefined.

So, for these principles, we might construct the following items to assess Properties.

- Without actually dividing, which result is larger? Explain how you know.
  - $n \div \frac{1}{2}$
  - $n \div \frac{1}{3}$
  - $n \div \frac{3}{4}$

- Describe what happens to the result of $\frac{3}{4} \div n$ as n becomes a larger and larger number.

Now think about the dimension of Uses. Do students have a good sense of when fraction division is needed? Can students use fraction division appropriately to solve application problems? Can students create their own problems where fraction division might be needed? These questions might lead to the following assessment items in the dimension of Uses.

- Balpreet has $1\frac{2}{3}$ cups of rice. She uses $\frac{1}{6}$ cup for each serving. How many servings does she have?
- Make up a real-world problem for which $1\frac{2}{3} \div \frac{1}{6}$ would be used to solve the problem.

Finally, think about the dimension of Representations. What are the visual images related to fraction division that you want students to have? Should they be able to illustrate a division problem with a diagram? Should they be able to use manipulatives to illustrate the computation? These ideas provide a basis for constructing assessment items in the dimension of Representations.

- Draw a diagram to illustrate the meaning of $1\frac{2}{3} \div \frac{1}{6}$.

Each of the items illustrated in this section gets at a different perspective of students' understanding of fraction division. Although skills are important, students also need to understand the properties, uses, and representations of the concepts to have the knowledge to think flexibly about fraction division. Together, students' achievement on the various items can help teachers determine what aspects of the concept might need to be a point of focus during lessons.

3.2 *Examples of SPUR at the secondary level*

In this section, we consider how SPUR might be used to develop assessment items for the algebraic topic of solving linear equations. Although algebra is a gateway to much further study in mathematics, students need more than a manipulative facility with solving linear equations to have the background to study later topics related to functions or calculus.

Consider a typical skill item related to solving linear equations:

- Solve for x: $3x + 12 = 5x$.

Teachers know that many students who are able to solve simple equations such as this one have difficulty when the equation becomes more complicated. Perhaps they guessed at a solution or obtained a solution by trial and error. Although such approaches are appropriate in many situations, as students progress in mathematics they need to understand the equation-solving process. So, what principles might teachers want to ensure that students know relative to this topic? Some principles we would list include: (1) know that a number can be added to both sides of an equation to yield an equivalent equation; (2) know that both sides of an equation can be multiplied or divided by a non-zero number to yield an equivalent equation; (3) a solution to an equation must yield the same value for both sides of the equation. These principles form the basis for potential assessment items in the dimension of Properties.

- In solving $3x + 12 = 5x$, Yiping wrote $12 = 8x$ as the next step. Is Yiping correct or not? How do you know?
- To solve $3x + 12 = 5x$, Muhundan wrote the following steps:
  Step 1: $3x + 12 = 5x$
  Step 2: $12 = 2x$
  Step 3: $6 = x$
  Explain what Muhundan did to go from each step to the next.

Many students wonder when they will ever use algebra in the real world. So, it is important that students recognize application problems that can be solved by using an equation. In addition to solving application problems created by the teacher, students should be able to generate their own application problems as well. Both ideas provide a basis for creating items that assess the dimension of Uses.

- Make up a real-world problem that can be answered by solving $3x + 12 = 5x$. Be sure to specify the meaning of the variable.
- Two children are saving their money to buy a special video game. Carlos already has 12 dollars and saves \$3 each week. Andrew does not have any money already saved but decides to save \$5 each week. If neither one takes out any of their savings, in how many weeks will they have the same amount saved?

Both problems give teachers insight into how their students address application problems involving linear equations. We have found that many students who are able to solve an equation have difficulty writing a realistic real-world problem. So, teaching might focus on helping students think about their own situations that lead to linear equations.

Finally, what representations might teachers want students to have related to solving linear equations? Should students be able to solve the equation with a visual? Should they be able to create a table of values or two graphs to solve the equation? These ideas suggest the following item related to the dimension of Representations.

- Use the table of values to find a solution to $3x + 12 = 5x$.

x	$3x + 12$	$5x$
0	12	0
1	15	5
2	18	10
3	21	15
4	24	20
5	27	25
6	30	30
7	33	35
8	36	40

As with the sequence of items assessing fraction division, we would argue that each item gives insight into important aspects of solving linear equations. For instance, students who understand how to use a table to solve an equation can use this approach to solve equations in later years, such as $\sin x = e^x$, for which no algorithmic approach exists. As teachers consider which aspects of equation solving their students have mastered and which they have not, they can design or modify teaching to help students develop a well-rounded and balanced view of the topic.

4 A Look at Achievement in Terms of SPUR

If teaching has focused on a multi-dimensional perspective, then SPUR provides a lens through which to consider student achievement as well. The International Project on Mathematical Attainment (IPMA) was a multi-country longitudinal study that tracked students' growth in mathematics from their first year of schooling through the end of primary schooling, generally 5 to 6 years. Both Singapore and the United States participated in this project. In Singapore, students from three different schools participated, with 856 remaining in the study from primary one (age 6) through primary five (age 10) (Kaur, Koay, and Yap, 2004). In the United States, students from six schools in two different states participated, with 181 students remaining in the study from kindergarten (age 5) through grade 5 (age 10) (Thompson, 2004).

Test 6, given to students in their respective countries at the end of primary schooling, consisted of 140 items. Overall, the mean percents correct (mean percent correct is the mean score of the cohort as a percentage of total possible score for the test) for students from the United States and Singapore were 71% (standard deviation = 11%) and 76% (standard deviation = 12%), respectively. These overall results suggest that students in Singapore scored somewhat higher than those in the United States. However, these overall results provide little insight into what might account for any differences. Did Singaporean students score consistently higher than students from the United States? Were there some dimensions for which students in the United States performed better? For teachers who want to modify their teaching to raise student achievement, where should they focus their teaching? The lack of specific information suggests a need to take a closer look at the achievement results.

Analysis of the 140 items of Test 6 showed that there were 56 (40%) items dealing with Skills, 25 (18%) items dealing with Properties, 17 (12%) items dealing with Uses, and 42 (30%) items dealing with Representations. One may note that this test was not designed to measure growth in a student's attainment according to the dimensions of understanding (SPUR). Hence, it is inevitable that a post-construction analysis of the test has yielded varying proportions of items in the four dimensions, Skills, Properties, Uses and Representations. Figure 2 shows the achievement for students in each county by the dimensions of understanding.

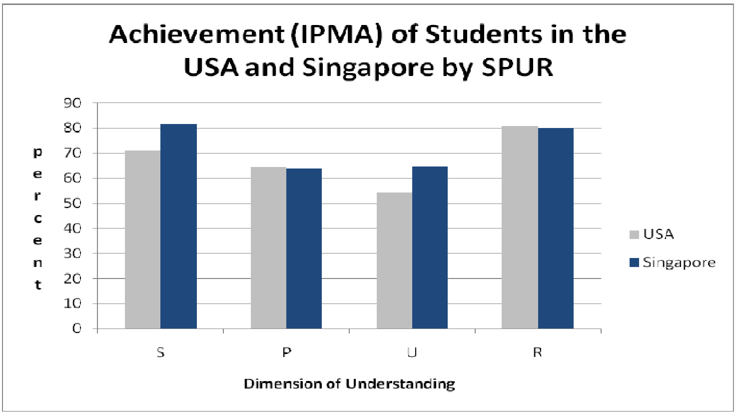


Figure 2. Comparison of achievement (IPMA) for USA and Singapore students by dimension of understanding

From Figure 2, several observations about the results are apparent. First, achievement across the four dimensions was not consistent within a country. Second, in both countries, students did better on Skills and Representations than on Properties or Uses. In both countries, achievement on Uses was about 15% lower than on Skills. Third, United States and Singapore students performed comparably on items dealing with Properties and Representations. The differences in achievement between students in the two countries were due primarily to differences in achievement on Skills and on Uses.

In addition to analyzing the results by SPUR, the items on this test were also analyzed by content. Among the 140 items, Table 1 shows how the SPUR items were distributed among five main content categories. As evident in the table, although the content area of Number was assessed from all four dimensions, this was not the case for the other content areas. So, teachers did not have an opportunity to determine how their students might address some aspects of the other content areas.

Table 1

Number of items by Content and SPUR for Test 6 of IPMA

	Number	Algebra	Measure	Geometry	Data/Chance	Total
Skills	40	3	7	0	6	56
Properties	16	0	0	8	1	25
Uses	9	1	6	0	1	17
Representations	22	1	6	7	6	42
Total	87	5	19	15	14	140

Figure 3 shows the achievement by United States and Singaporean students by the dimension of understanding for the four content areas, provided there were at least five items in the related cell. The picture of this analysis also provides evidence that teachers can use to modify and enhance their teaching. For instance, for Singaporean students, achievement within Number was uniformly high across all four dimensions; however, in Measure, achievement was high for Skills and Representations but low for Uses. In Data/Chance, achievement in Skills was considerably below the achievement in Representations. For Geometry, achievement in both of the assessed dimensions was low. The fact that several dimensions of understanding were not even assessed in Geometry and in Data/Chance suggests that teachers' pictures of their students' knowledge was not complete.

For students in the United States, their achievement in Representations was higher in each content area than any of the other dimensions. Although the achievement in the four dimensions for Number was somewhat consistent, the difference in achievement across dimension for the other content areas suggests teachers have considerable work to help students develop a more robust understanding of mathematics.

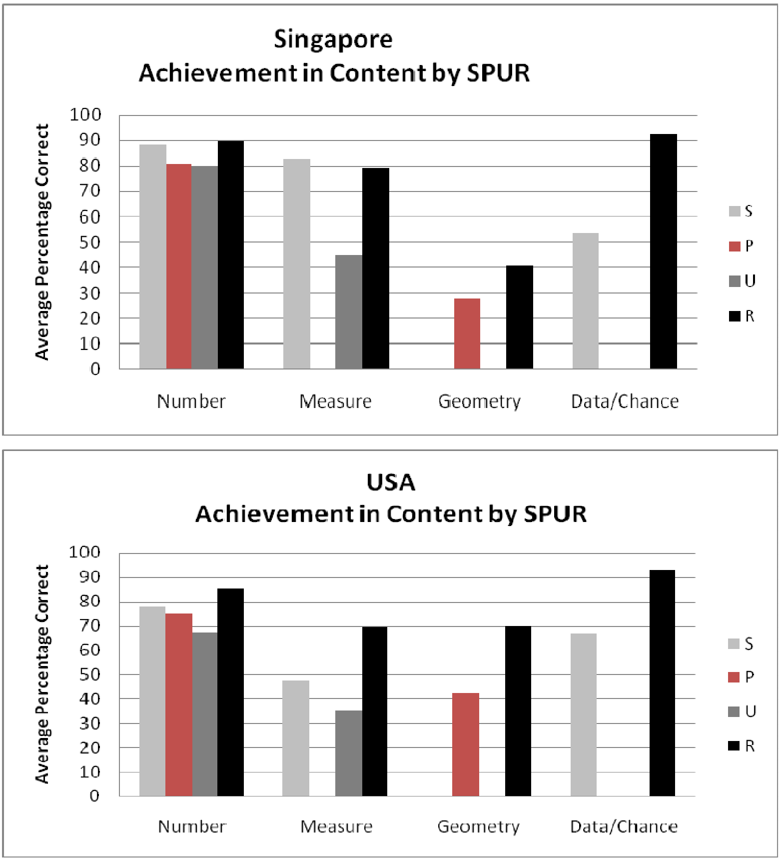


Figure 3. Comparison of achievement (IPMA) by Content and SPUR for students in the U.S. and Singapore

5 Discussion and Conclusion

The comparison of test results from students in the United States and in Singapore suggests that much can be learned by using a more balanced approach to assessment, such as that suggested by SPUR. Overall test results provide only a quick view of student understanding, and a view that can be misleading. The analysis of test results by dimension of understanding suggests that students' abilities are not consistent across skills, properties, uses, and representations. In addition, their achievement across the four dimensions also varies by the content area being assessed.

Teachers typically use assessments to determine what their students have mastered and to determine strengths and weaknesses of teaching. Analysis of test results by SPUR can help teachers target their classroom activities to those aspects of the content that students have not yet mastered so that students develop a robust and balanced understanding of mathematics. Teachers can determine whether (1) they focused their teaching on all four dimensions but students failed to achieve, or (2) they focused their teaching on only some dimensions so teaching needs to be modified to incorporate other important facets of mathematics (Bleiler and Thompson, 2010).

If our instructional goal is to develop students with a robust and flexible understanding of mathematics, then it is essential that we assess more than just their knowledge of skills. Although it is relatively easy to write skill items, with knowledge of the other dimensions of understanding teachers can modify those skill items or write entirely new items to gain additional perspectives on their students' understanding of the concept. Insights of students' understanding of concepts can be used to modify teaching so students build a solid foundation of mathematical knowledge.

Although we have suggested SPUR as a tool to ensure a balanced view of mathematical understanding, as previously indicated there are other models for considering multiple perspectives to learning (e.g., Freudenthal, 1983; National Research Council, 2001). What is crucial is that assessment and teaching align. If teaching incorporates a multi-dimensional view of understanding, then assessment also needs to incorporate such a perspective.

References

Bleiler, S. K., & Thompson, D. R. (2010). Dimensions of mathematics understanding: A longitudinal analysis of primary students' achievement in the United States. In Y. Shimizu, Y. Sekiguchi & K. Hino (Eds.), *Proceedings of the 5th East Asia Regional Conference in Mathematics Education,* Volume 2 (pp. 400-407). Japan, Tokyo: Japan Society of Mathematical Education.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht, The Netherlands: Reidel.

Kaur, B., Koay, P. L., & Yap, S. F. (2004). Country reports: Singapore. In D. Burghes, R. Geach, & M. Roddick (Eds.), *International Project on Mathematical Attainment Report. Series of international monographs on mathematics teaching worldwide*, Monograph 4 (pp. 175-190). Hungary: Müszaki Könyvkiadó, a WoltersKluwer Company.

Krutetskii, V. (1976). *The psychology of mathematical abilities in school children*. (translated by J. Teller. Edited by J. Kilpatrick and I. Wirszup). Chicago, IL: University of Chicago Press.

Ministry of Education. (2006a). *Mathematics syllabus – Primary*. Singapore: Author.

Ministry of Education. (2006b). *Mathematics syllabus – Secondary*. Singapore: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Thompson, D. R. (2004). Country reports: USA. In D. Burghes, R. Geach, & M. Roddick (Eds.), *International Project on Mathematical Attainment Report. Series of international monographs on mathematics teaching worldwide,* Monograph 4 (pp. 217-227). Hungary: Müszaki Könyvkiadó, a WoltersKluwer Company.

Thompson, D. R., & Senk, S. L. (2008, July). *A multi-dimensional approach to understanding in mathematics textbooks developed by UCSMP*. Paper presented in Discussion Group 17 of the International Congress on Mathematics Education. Monterrey, Mexico.

Usiskin, Z. (2003). A personal history of the UCSMP secondary school curriculum: 1960-1999. In Stanic, G. M. A., & Kilpatrick, J. (Eds.), *A history of school mathematics*, Volume 1 (pp. 673-736). Reston, VA: National Council of Teachers of Mathematics.

Usiskin, Z. (2007). The case of the University of Chicago School Mathematics Project: Secondary Component. In C. R. Hirsch (Ed.), *Perspectives on the design and development of school mathematics curricula* (pp. 173-182). Reston, VA: National Council of Teachers of Mathematics.

Chapter 3

Assessing Problem Solving in the Mathematics Curriculum: A New Approach

TOH Tin Lam QUEK Khiok Seng LEONG Yew Hoong
Jaguthsing DINDYAL TAY Eng Guan

In this chapter, we focus on the implementation of a framework for assessing problem solving in a specifically designed curriculum. While traditional assessment of problem solving has focused on the products of problem solving, this framework builds on the works of Pólya and Schoenfeld and gives greater emphasis to the processes. This assessment framework works in tandem with a practical worksheet which is an important feature of the problem-solving curriculum that we have designed. We present the assessment framework and how it is used to assess students' doing problem solving. In particular, we use the assessment framework to assess the works of two students, Zill and William. We also discuss the students' ideas about the problem-solving curriculum and the assessment framework.

1 Introduction

It is generally accepted that the right processes will lead to a good product. In mathematical learning, processes are often assessed indirectly, i.e., by assessing the products as it is almost impossible to access processes directly. Therefore, it comes as no surprise that overwhelmingly the assessment of problem solving has focused on assessing the products of the problem-solving process. However, assessing only the products of the learning process is no guarantee that correct processes have been followed. In this chapter, we look at the assessment of problem solving from a new perspective in a specifically designed problem-solving curriculum that amply focuses on the processes without neglecting the products of the learning process. We present a worksheet, called the Practical Worksheet, alongside a scoring rubric that gives credit to students' thinking throughout the problem-solving process. We highlight, with appropriate interviews, the use of the rubric in grading the work of two students and illustrate how the use of the rubric together with the worksheet helped students with their metacognition when they engaged in problem solving.

There has been much interest in mathematical problem solving since Pólya published his first book on mathematical problem solving (Pólya, 1945). From the 1980s onwards, there has also been a world-wide push for problem solving to be the central focus of the school mathematics curriculum. For example, in the United States, the National Council of Teachers of Mathematics (NCTM) in their document on the principles and standards for school mathematics stated that "[p]roblem solving should be the central focus of the mathematics curriculum" (NCTM, 2000, p. 52). Mathematical problem solving has been at the heart of the Singapore mathematics curriculum since the 1990's. The stated primary aim of the Singapore mathematics curriculum is to develop students' ability to solve mathematics problems (Ministry of Education, 2006). The centrality of mathematical problem solving is clearly depicted in the framework of the Singapore mathematics curriculum, as shown in Figure 1.

Implementing a problem-solving curriculum is quite challenging. First, there is the issue of the meanings attached to the term "problem solving". Teachers in Singapore vary in their styles and approaches to problem-solving instruction in terms of the amount of class time and attention spent in the various classroom activities, and differing emphasis on each of the four problem stages of Pólya (Hedberg, Wong, Ho, Lioe and Tiong, 2005). Second, a major hurdle in the Singapore context is the routinisation of certain aspects of problem solving, such as the teaching of problem-solving heuristics that match certain problem types. Also, the

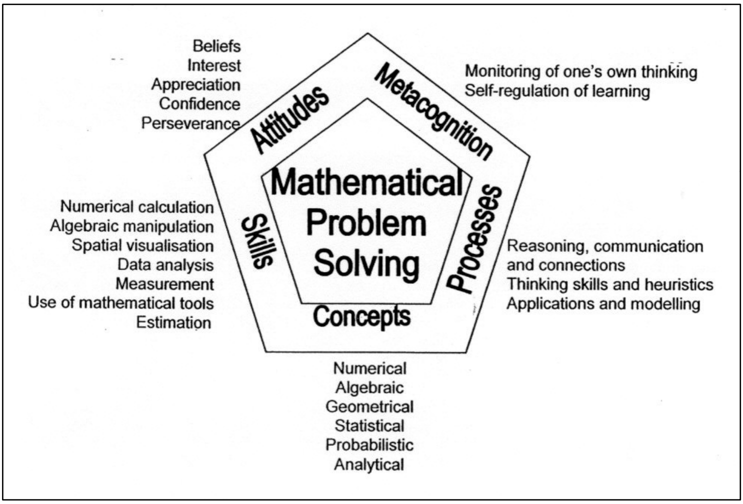


Figure 1. Framework of the school mathematics curriculum

over-emphasis during classroom teaching about the types of mathematical problems which are usually found in high-stakes national examinations adds to this issue, somewhat similar to what happened in New Zealand as reported by Holton, Anderson, Thomas and Fletcher (1999). The New Zealand Ministry of Education developed a national numeracy project which emphasized a problem-solving approach and it has now been introduced to the majority of primary schools in that country. However, success is so far limited to the primary level (Ministry of Education New Zealand, 2006), as high-stakes examinations have blunted the problem-solving approach in mathematics classes at the secondary level.

As mentioned earlier, traditionally, the assessment of problem solving in the classroom has focused on assessing the products rather than the processes of problem solving. Our efforts to meet the challenge of teaching mathematical problem solving to students call for a curriculum that emphasizes the processes (while not neglecting the products) of problem solving and an assessment strategy to match it so as to drive the mode of teaching and learning of mathematics.

2 Mathematical Problem-Solving Model

Good problem solvers presumably have built up their own models of problem solving. Having a model of problem solving is especially important when an individual's progress in solving a mathematical problem is not smooth. A problem-solving model that is made explicit to students should be helpful in guiding them in the learning of problem solving, and in regulating their problem solving attempts. Even a good problem solver may find the structured approach of a model useful. As Alan Schoenfeld (1985) recounts in the preface to his book *Mathematical Problem Solving* about Pólya's book *How to Solve It*:

> In the fall of 1974 I ran across George Pólya's little volume, *How to Solve It*. I was a practising mathematician … My first reaction to the book was sheer pleasure. If, after all, I had discovered for myself the problem-solving strategies described by an eminent mathematician, then I must be an honest-to-goodness mathematician myself! After a while, however, the pleasure gave way to annoyance. These kinds of strategies had not been mentioned at any time during my academic career. Why wasn't I given the book when I was a freshman, to save me the trouble of discovering the strategies on my own? (p. xi)

The practical approach which we describe later uses Pólya's model as the basis which we have enhanced with Schoenfeld's (1985) ideas about problem solving. Pólya's model is well-known and it is mentioned in the syllabus document of the Singapore Ministry of Education. We wanted a model which is most familiar to those who have to work within the Singapore mathematics syllabus. We remark that any other sensible model of problem-solving would be equally useful (see for example, Mason, Burton and Stacey, 1985).

The essential features of Pólya's problem-solving model are shown in Figure 2.

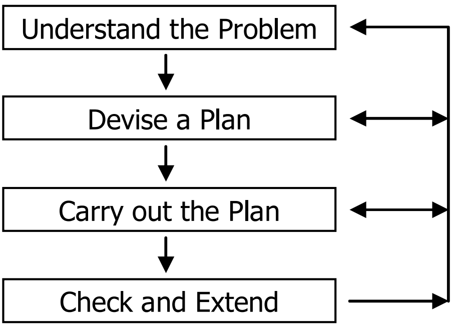


Figure 2. Flowchart of Pólya's problem-solving model

The model is depicted as a flowchart with four components, *Understand the Problem*, *Devise a Plan*, *Carry out the Plan*, and *Check and Extend*, with back-flow allowed to reflect the dynamic and cyclical nature of problem solving (Carlson and Bloom, 2005).

Schoenfeld (1985) grappled with the apparent worth of Pólya's model and the real-world failure of its application in the classroom. He argued that successful problem solving required more than just a direct application of the model; other factors are crucial as well. His research culminated in the construction of a framework for the analysis of complex problem-solving behaviour. The four aspects highlighted in his framework are:

- Cognitive resources – the body of facts and procedures at one's disposal
- Heuristics – 'rules of thumb' for making progress in difficult situations
- Control – having to do with the efficiency with which individuals utilise the knowledge at their disposal
- Belief systems – one's perspectives regarding the nature of a discipline and how one goes about working on it

Anecdotal evidence from mathematics classrooms in Singapore shows that the teaching of problem solving in schools has typically emphasized the teaching of heuristics (Hedberg et al., 2005). The right choice and use of heuristics were assumed to be sufficient for successful problem solving. Schoenfeld's framework suggests that we need to provide students with more than just Pólya's model and a range of heuristics. The students have to manage resources at their disposal, choose promising heuristics to try, control the problem-solving process and progress, examine their beliefs about mathematics that hinder or facilitate problem solving, and in culmination, generalise and extend.

3 Mathematics Practical – A New Paradigm

Our combined classroom experiences of teaching problem solving strongly suggest to us that students are generally resistant to apply the stages of Pólya's model. They also do not consciously use and manage heuristics productively. Even the higher achieving students who could solve the given problems do not generally make the extra effort to finally check and extend the problem.

In an attempt to help students cultivate the discipline of good problem-solving habits (as explicated by Pólya and Schoenfeld), especially when they are clearly struggling with the problem, we decided to construct a worksheet like that used in science practical lessons and told the students to treat the problem-solving lesson as a mathematics "practical" lesson. In this way, we hope to achieve a paradigm shift in the way students look at these "difficult, unrelated" problems which had to be done in this "special" classroom setting – a form of mathematics "practical".

The use of practical work to achieve the learning of the scientific processes has a long history of at least a hundred years and can be traced to Henry Edward Armstrong (Woolnough and Allsop, 1985). Woolnough and Allsop (1985) stated clearly what is to be achieved in science education:

> As we look at the nature of science we see two quite distinct strands. The knowledge, the important content and concepts of science and their interrelationships, and also the processes which a scientist uses in his working life. In teaching science we should be concerned both with introducing students to the important body of scientific knowledge, that they might understand and

> enjoy it, and also with familiarizing students with the way a problem-solving scientist works, that they too might develop such habits and use them in their own lives. (p.32)

It is instructive to see that we could just replace ‘science’ with ‘mathematics’ and the preceding passage reads just as true, which any mathematics educator would agree. It is certainly conceivable that similar specialised lessons and materials for mathematics may be necessary to teach the mathematical processes, including and via problem solving (Toh, Quek and Tay, 2008). This approach could elicit the learning of the processes of problem solving, analogous to the processes of science practical skills of scientists in their working life.

4 Mathematics Practical Worksheet

Tay, Quek, Toh, Dong and Ho (2007) introduced the “mathematical practical” into problem-solving lessons using a “practical” worksheet. The students were encouraged to treat the problem-solving class as a mathematics “practical” lesson.

The worksheet contains sections explicitly guiding the students to use Pólya’s stages and problem-solving heuristics to solve a mathematics problem. A complete practical worksheet is given in Appendix A.

5 Mathematics Practical Lessons

A problem-solving lesson, consisting of 55 minutes, is divided into two parts. In the first part (except for Lesson 1), the teacher reviews homework of the last lesson and explains one aspect of problem solving, such as a stage in Pólya’s model. The second part focuses on one problem, the ‘Problem of the Day’. Thus, the entire class focuses on only one problem during each lesson. Typically, the problem-solving module requires 10 lessons. Table 1 shows the outline of the problem-solving module.

Each page of the worksheet (see Appendix A) corresponds to one stage of Pólya's problem-solving model. The use of the worksheet is gradually introduced across a few lessons. In Lesson 2 when Polya's model is explicitly explained, the students will work on the Problem of the Day on a modified worksheet. This has only the first page (Understand the Problem) to be filled in, while the usual work is carried out on blank pieces of paper, as Lesson 2 emphasizes the importance of understanding a given problem. In the next two lessons where the emphasis is on using heuristics, the modified worksheet will include Pages 1 to 3, and the students will have to fill in their work for Understand the Problem and Devise a Plan, and carry out the usual working in the Carry out the Plan page(s). In the fifth lesson, the Practical Paradigm and Practical Worksheet are explained and the full worksheet is implemented.

When solving problems, ideally, the student will follow the model and go through all the four stages, with suitable loopbacks (to be worked on blank pieces of paper and attached at the relevant place). However, so as not to straitjacket an unwilling problem solver, the student may jump straight to Stage 3 (Carry out the Plan) and is given up to 15 minutes to solve the problem. Students who complete the problem in time need only do the Stage 4 of Pólya's model. Students who fail to make any progress in 15 minutes will be required to go through all four stages of Pólya's model. This is so, as the rationale is that going through all the stages systematically would encourage the unsuccessful student to meta-cognate and deliberate on the solution process, eventually producing a better solution. The successful student is allowed to leapfrog to Stage 4, which is important for consolidation of the method and gaining a fuller insight to the problem. Certainly, a particular student may have to go through all the four stages for one problem and not for another problem, where the plan to him is obvious and he needs only to Check and Extend. The choice allowed here is to show that explicit use of Pólya's model is very useful when one is stuck with a particular problem.

Table 1
The 10 lesson problem-solving module

Lesson	Activity
1	• Distinguish between a problem and exercise • Model successful problem solving
2	• Introduce Polya's problem solving model • Introduce Stage I of the Practical Worksheet (Understand the Problem)
3	• Introduce the meaning of the word *heuristics* and provide a list of the common heuristics for mathematical problem solving • Introduce Stages I to III of the Practical Worksheet (Understand the Problem, Devise a Plan, Carry out the Plan)
4	• More on heuristics • Practice on using Stages I to III of the Practical Worksheet
5	• Introduce to the practical paradigm of mathematical problem solving • Formal use of the Practical Worksheet to solve Problem of the Day and Homework Problem
6	• Focus on Check and Extend, i.e., Stage IV of the Practical Worksheet • Emphasis on adapt, extend and generalize a mathematical problem • Introduce the assessment rubric
7	• Identify the features of Check and Extend
8	• Introduce the importance and use of Control (Schoenfeld, 1982) in mathematical problem
9	• Introduce the use of the Control Column in Stage III of the Practical Worksheet
10	• Revision on the key features and processes of mathematical problem solving

6 The Scoring Rubric

It is common knowledge among teachers and educators that most students will study mainly for curricular components which are to be assessed. There needs to be a corresponding assessment strategy that drives the teaching and learning of problem solving as described in the preceding paragraphs. Effective assessment practice "begins with and enacts a vision of the kinds of learning we most value for students and strive to help them achieve" (Walvoord and Anderson, 1998). To assess the students' problem-solving processes (which we value), we developed a scoring rubric based on Pólya's model and Schoenfeld's framework.

The scoring rubric focuses on the problem-solving processes highlighted in the Practical Worksheet. There are four main components to the rubric, each of which would draw the students' (and teachers') attention to the crucial aspects of, as authentic as possible, an attempt to solve a mathematical problem. In establishing the criteria for each of these components of problem solving, we ask the question, What must students do or show to suggest that (a) they have used Pólya's approach to solve the given mathematics problems, (b) they have made use of heuristics, (c) they have exhibited "control" over the problem-solving process, and (d) they have checked the solution and extended the problem solved (learnt from it)?

The rubric is outlined below. The complete rubric is attached in Appendix B.

- *Pólya's Stages* [0-7 marks] – this criterion looks for evidence of the use of cycles of Pólya's stages (Understand the Problem, Devise a Plan, Carry out the Plan), and correct solutions.
- *Heuristics* [0-7 marks] – this criterion looks for evidence of the application of heuristics to understand the problem, and to devise/carry out plans.
- *Checking and Extending* [0-6 marks] – this criterion is further divided into three sub-criteria:
    - Evidence of checking of correctness of solution [1 mark]
    - Providing for alternative solutions [2 marks]
    - Extending and generalizing the problem [3 marks] – full marks for this is awarded for one who is able to provide (a) two or more problems with solutions or suggestions to solution, or (b) one significant related problem with comments on its solvability.

The rubric was designed to encourage students to go through Pólya stages when they are faced with a problem, and to use heuristics to explore the problem and devise a plan. They would return to one of the first three stages (see Practical Worksheet) upon failure to realize a plan of solution. Students who show control (Schoenfeld's framework) over the problem-solving process gain marks. For example, a student who did not manage to obtain a completely correct solution would be able to score up to eight and three marks each for *Pólya's Stages* and for *Heuristics*, making a total of eleven, if they show evidence of cycling through the stages, use of heuristics, and exercise of control.

The rubric allows the students to score as many as 70% of the total 20 marks for a correct solution. However, this falls short of obtaining a distinction (75%) for the problem. The rest would come from the marks in *Checking and Extending*. Our intention is to push students to check and extend the problem (Stage 4 of Pólya's stages), an area of instruction in problem solving that has not been largely successful so far (see for example, Silver, Ghousseini, Gosen, Charalambous, and Strawhun, 2005).

7 Students' Responses and Assessment

The researchers conducted the first series of ten mathematics practical lessons as an elective module to a group of 24 Year 8 students (age 14) in a Singapore secondary school. The students' use of the practical worksheet in answering the questions was indicative of their method of solution, their errors and misconception related to each question. Both the school teachers and researchers used the scoring rubric to mark the students' responses to the "Problem of the Day" and problem for homework. The inter-rater reliability (κ) was 0.695. This was acceptable because the teachers were not yet experienced in marking using the rubric nor with the solutions and extensions to the problems. Overall, after discussion, the marking by both groups was largely consistent for most of the problems.

The students were informed of the criteria of scoring of their assignments, which was done after the sixth lesson. It was deemed that the students needed time to get accustomed to the use of the practical worksheet and the scoring.

Qualitative information was obtained from interview sessions with selected students. The followings were some of the prompts used by the interviewer for the student interviews:

- Why did you sign up for the (mathematics problem-solving practical) course?
- Name one thing that you learnt from the course.
- How does the practical worksheet help you in solving problems?
- What do you think of the assessment of the course?

In this section, we shall present the interview segments with two of the students, whom we coded as William and Zill (pseudonyms). William was the highest achieving student in mathematics who also represented the school in various National Mathematics Competitions. Zill was a 'non-believer' in mathematical problem solving. We use W, Z and R to denote William, Zill, and the Researcher respectively. Relevant segments of the interviews with our interpretation alongside them are presented below. We will also show how Zill's and William's problem-solving attempts, during the final test, were graded with the help of the assessment rubric.

	Interview with Zill	***Interpretation***
Z:	It's not that worth it of my time cause I didn't really learn much other than solving problems and getting exposed to new problem. I didn't learn much. I expected something even more. My expectations were higher.	
R:	Like what?	
Z:	Like maybe… possible… real like problems and not just like solving, solve, simplify this equation or prove that. As in something that's of more real life problem solving. That's what I expected when I saw the topic.	*Zill had expected this course to be one involving applications of mathematics in real life applications instead of one focussing on mathematical problem solving.*
R:	Can you tell us one thing you have learnt?	
Z:	Hardly any.	
R:	Which part of the course you find is hard to handle?	
Z:	Hardest to handle would be trying to answer some of the questions that I've no answer to like which method did you use like the process of problem solving. For example, like devise a plan. What are the	*To Zill, mathematical problem solving is about obtaining the correct mathematical solution to a problem and not following explicitly the processes of*

	key concerns and right now you're a chain of thinking I mean when I think how to solve a problem I just go to the flow and I just think. I don't really care like what I think. As long as I can solve it I just solve it. And making me like try to write out what I'm thinking exactly and everything that goes in the brain is like very abstract so it's very hard to put it into words….	*problem solving and committing them to words.*
Z:	Problem solving is like you just get the answer and the problem is solved already.	
Z:	Sounds practical is more like handling physical thing. Practical as like what I expected. Outside practical is like you go measure the … then calculate the… ya all these… this is more of practical. This is more like theory. The theory before binary numbers all these etc.	*Zill felt that the use of the word "practical" was not so appropriate for this course, as this course is more on "theory".*
R:	Ok. So you see practical more as a hands-on, right?	
Z:	Ya..This is thinking la. Thinking then is theory.	
Z:	I don't know because a lot of my work was late and one of them never write my name so if you… Oh but the Dr. Tay [i.e. the researcher] said that if you're late once, then you count in all the assignments. Was he serious?	*Zill was concerned with his final grade he would obtain for this course.* *(Initially Dr Tay mentioned that the grade would be the average of the three best assignments if all the assignments were handed in.)*
Z:	I use it. Ya. Certain heuristics.	
R:	So the course is not totally useless to you?	
Z:	Ya but I knew those heuristics before so ya it's still useful but ya I didn't learn something new from this course.	*Zill indicated eventually that he learnt problem solving heuristics which were actually helpful to him.*

R: Oh. Ok. Ok. So you did find something useful. So what in mind regard to the content of the course you think to make this course more useful to you. What you think should have been included into the content?

Z: Like something new that we haven't learn. Something new that we haven't learn and not like this algebra we can use this and solve a more practical problem like find a… devise a formula for something. More practical la… meaning more hands-on

Zill felt that more practical questions could have been used in the course.

Figures 3 and 4 show Zill's attempt, in the final test, to solve the given problem.

Problem

There are two timers: one for 5 minutes and one for 9 minutes. We want to heat a beaker of water for exactly 11 minutes. How can we do this using only these timers?

Instructions

- You may proceed to complete the worksheet doing stages I – IV.
- If you wish, you have 15 minutes to solve the problem without explicitly using Polya's model. Do your work in the space for Stage III.
  - ➢ If you are stuck after 15 minutes, use Polya's model and complete all the stages I – IV.
  - ➢ If you can solve the problem, you must proceed to do stage IV – Check and Extend.

I Understand the problem

(You may have to return to this section a few times. Number each attempt to understand the problem accordingly as Attempt 1, Attempt 2, etc.)

(a) Write down your feelings about the problem. Does it bore you? scare you? challenge you?

(b) Write down the parts you do not understand or that you misunderstood.

(c) Write down the heuristics you used to understand the problem.

Attempt 1

a) It surprises me because it is not a problem

b) I understand it completely

c) none

II Devise a plan

(You may have to return to this section a few times. Number each new plan accordingly as Plan 1, Plan 2, etc.)
(a) Write down the key concepts that might be involved in solving the question.
(b) Do you think you have the required resources to implement the plan?
(c) Write out each plan concisely and clearly.

Plan 1

a) algebra

b) Yes

c) we need to express 11 minutes in the form of sum and differences of 5 and 9

Figure 3. Zill's solution - Stages I and II

III Carry out the plan

(You may have to return to this section a few times. Number each implementation accordingly as Plan 1, Plan 2, etc., or even Plan 1.1, Plan 1.2, etc. if there are two or more attempts using Plan 1.)
(i) Write down in the *Control* column, the key points where you make a decision or observation, for eg., go back to check, try something else, look for resources, or totally abandon the plan.
(ii) Write out each implementation in detail under the *Detailed Mathematical Steps* column.

Detailed Mathematical Steps	*Control*
Attempt 1	
$11 = 5+5+1$ $= 5+5+(5+5-9)$ $= -9+5+5+5+5$	Express 11 as sum and difference of 5 and 9
time before heating starts − means ~~to start the timer~~ before, and + means ~~to start the timer~~ after.	Find out what they mean
Steps: - start the 9 minute and 5 min. timer. - When 5 min timer ~~runs out~~ rings, restart it; 9 min. timer is now at 5th minute. - When 9 min timer ~~runs out~~ rings, start heating: ~~5~~ 5 min. timer is now at 4th minute. - 1 minute has passed when 5 min. timer rings again. - ~~when last~~ Restart the 5 minute 2 more times (i.e. add 10 more minutes) - water heated successfully after second ring! Yay!	

3

IV Check and Extend

(a) Write down how you checked your solution.
(b) Write down your level of satisfaction with your solution. Write down a sketch of any alternative solution(s) that you can think of.
(c) Give one or two adaptations, extensions or generalisations of the problem. Explain succinctly whether your solution structure will work on them.

a) By staring at the classroom clock and tapping my stationary to represent a ring from the timer

b) Quite satisfied

c) Let m & n be duration of timers, -

for them to be able to time any time,

$mx - ny = 1$, where $x > 0$ and $y > 0$, as 1 is a factor of all numbers

x and y are any positive integer

Let p be the time ~~needed to time~~ ~~pee~~ wanted, and $\frac{p}{z}$ be a factor of it

if $mx - ny = \frac{p}{z}$,

$\Rightarrow mxz - nyz = p$

since x and y are arbituarily assigned,

4

For p to be able to be timed, $mx - ny = p$

Figure 4. Zill's solution - Stages III and IV.

Zill's solution was correct. In Stage III (Carry out the plan), he demonstrated his plan of expressing 11 as the sum and difference of 5's and 9's. He had a clear plan of tackling the problem. Hence, under Pólya's stage, he was awarded 10 marks (Level 3 of correct solution). He demonstrated the use of heuristics in Stage II (Devise a plan) and Stage III (Carry out the plan). Furthermore in Stage III, he listed the steps clearly on how 11 minutes can be obtained. Thus, under Heuristics, he was awarded 4 marks (Level 2 of a correct solution). Under Stage IV (Check and extend), he did not demonstrate effort to check the reasonableness of his solution or attempt to provide an alternative solution to the problem. However, he offered a possible generalization of the given problem, which was a problem involving Diophantine equations. His total score based on the scoring rubric was 15 out of 20. The detailed breakdown of his scoring is shown in Table 2 below.

Table 2

Score for Zill

	Descriptors	**Marks awarded**
Pólya's stages	Correct Solution – Level 3	10
Heuristics	Correct Solution – Level 2	4
Check and Extend	Checking – Level 2	0
	Alternative Solution – Level 2	0
	Extend, Adapt & Generalize – Level 2	1
	Total Score for Zill	15

	Interview with William	***Interpretation***
R:	Researcher got William to talk about his interest in mathematics and training sessions in Mathematics Olympiad. Interesting starting point to discuss then. How is this in your mind when you have so many Olympiads training already? And then you attend Dr. Tay's problem solving course. What is your first impression and how are they similar and how are they different from Olympiad training?	*William was very interested in mathematics. He was high achieving in mathematics and was participating in weekly Olympiad training in his school.*
W:	This module focus mainly on how you're gonna start when solve a problem. How to become more efficient and at least you define your thinking process along the way so you won't get very confused in case something goes wrong.	*William found problem solving lessons complementing his Olympiad training in terms of refining his thinking processes.*

R: You said define your thinking process. Can you elaborate more what do you mean by define your thinking process?

W: Because it's like… before this module, when I saw a question, all the formulas will just rush into your mind. And then I'll start writing down… most of them in my head la because sometimes the brain is just faster than the hand so you cannot write everything down. And then a lot of careless mistake and stuff. After this module, it's like you think like step by step already, so you'll know when… you can like stop the formulas from rushing in at first, so think about what you're supposed to do everything, then the formula will slowly come in.

Going through the problem solving processes had helped William to be more careful in monitoring his thoughts.

W: Olympiad training is they just tell you the solution, you're supposed to learn from the solution and then apply the solution to other questions, whereas this type is they show you the thinking process of how the solution is derived if you don't have the solution and you're only given the question, and then solve it la.

William clearly distinguished between Olympiad training and the problem solving lessons.

R: Which part of the module that Dr. Tay taught you, you like the most that comes to your mind, and you really find that you like this?

W: It's like at first question, mostly I think there's only one solution, but as we discuss more and more, you can see a lot of other alternate solutions, which sometimes may be more efficient than your usual way of solving. Because sometimes an alternative solution can hold better than usual solutions or in some other questions you cannot apply your normal solution, you have to apply an alternative solution or twist it some way to solve the question.

William liked the part of looking for alternative solutions most. He appeared to appreciate different approaches of solving the same problem.

R: These are all your stuff.

W:	This one does make me think because I was taught the Olympiad way so everything I do is this formula. The just shows me how to do without using the formula. Make me really think about how to solve it. This was a good question. This one is… I can never think we need to use binary to do this. All the weights and everything.	*Olympiad training equipped William to solve a problem with formulae while problem solving lessons allowed him to solve a problem without first beginning with a formula.*
R:	Name one thing you learnt from the course.	
W:	Checking la. Check and… Yes. It's quite new to me. Ya because extension… realized that there are actually more to this problem than this normal problem. Once we know how to solve an extension or generalization then it's more or less you fully understand the problem, no matter how intrinsic, there's a way to solve it.	*Check and Extend is what William considered to have benefitted from the problem solving lessons.*
R:	Does it feel very uncomfortable to you that aiya every time solve already must extend. Do you really like it?	
W:	It's just at home right, you know when you think about, my class is a time is very important. I really have to use my brain to really think of a good extension or really enjoy this problem fully and … is very hard. At home I can… because I have a lot of time, I know I have a lot of time so I won't be distracted by anything, so I can fully just understand and enjoy this problem, and then extend whatever I like.	*William indicated that he enjoyed extending a problem, although it is very difficult to do during class lessons as it is time consuming. Nevertheless, extension allows him to understand and enjoy the mathematics problems.*
R:	Ah… in this course one of the main thing that Dr. Tay was trying to help us learn was the use of practical worksheet. You know what is the practical worksheet? The thing that you are staring at now is the practical worksheet.	
W:	The last page [Check and Extend], they should just add one more paper here, because if I want alternative solution, I want extension generalization, everything	*It was clear from the interview that William was really excited with the Check and Extend part of the problem solving processes, that he*

in detail, I really need a lot of paper. Ya. Although… because there are some questions I've already known the solution to, so it's page one, two and three it's like just…Ya it's the part four that is exciting. [i.e. Check & Extend]

proposed adding one more page for the Part Four (Check and Extend).

Figures 5, 6 and 7 show William's attempt in the final test on the same problem.

Problem

There are two timers: one for 5 minutes and one for 9 minutes. We want to heat a beaker of water for exactly 11 minutes. How can we do this using only these timers?

Instructions

- You may proceed to complete the worksheet doing stages I – IV.
- If you wish, you have 15 minutes to solve the problem without explicitly using Polya's model. Do your work in the space for Stage III.
  - ➢ If you are stuck after 15 minutes, use Polya's model and complete all the stages I – IV.
  - ➢ If you can solve the problem, you must proceed to do stage IV – Check and Extend.

I Understand the problem

(You may have to return to this section a few times. Number each attempt to understand the problem accordingly as Attempt 1, Attempt 2, etc.)

(a) Write down your feelings about the problem. Does it bore you? scare you? challenge you?

(b) Write down the parts you do not understand or that you misunderstood.

(c) Write down the heuristics you used to understand the problem.

Attempt 1

a) I feel challenged, because I've done a similar question once, although it is simpler.

b) None

c) I noticed that 5, 9, 11 have no common factors, and I tried to do it mentally for a while to get the hang of it.

II Devise a plan

(You may have to return to this section a few times. Number each new plan accordingly as Plan 1, Plan 2, etc.)

(a) Write down the key concepts that might be involved in solving the question.

(b) Do you think you have the required resources to implement the plan?

(c) Write out each plan concisely and clearly.

Plan 1

1. Define variables (x and y)
2. Set up equations
3. Solve the equation
4. Relate back to problem
5. Offer solution

Figure 5. William's solution - Stages I and II

III Carry out the plan

(You may have to return to this section a few times. Number each implementation accordingly as Plan 1, Plan 2, etc., or even Plan 1.1, Plan 1.2, etc. if there are two or more attempts using Plan 1.)

(i) Write down in the *Control* column, the key points where you make a decision or observation, for eg., go back to check, try something else, look for resources, or totally abandon the plan.

(ii) Write out each implementation in detail under the *Detailed Mathematical Steps* column.

Detailed Mathematical Steps	*Control*
Attempt 1	
Let the number of times the 5 minute timer is used by x, and the 9 minute timer be y.	Define variables
Therefore, $5x+9y=11$	Setting up equation
By the Eucledian Algorithm, $9=5+4$, $5=4+1$	Solving the equation
$1=5-4$ $=5-(9-5)$ $=2(5)-9$	
$11=22(5)-11(9)$ $=4(5)-1(9)+18(5)-10(9)=4(5)-1(9)$	
Hence we have the result: $4(5)-1(9)=11$	
This means that the 5 minute timer is used 5 times, while the 9 minute timer is used only once.	Relating back to problem
*Refer to attached paper	

IV Check and Extend

(a) Write down how you checked your solution.
(b) Write down your level of satisfaction with your solution. Write down a sketch of any alternative solution(s) that you can think of.
(c) Give one or two adaptations, extensions or generalisations of the problem. Explain succinctly whether your solution structure will work on them.

Checking

If the water starts to be heated after the 9 minute timer, the 5 minute timer will have 1 minute left. Then the 5 minute timer is played 2 more times, totalling to 5+5+1 = 11 minutes.

Alternative Solution 1

In order to form 11, we have the following ways:

2+9
5+6
2(5)+1
3+8
4+7 } very low chance of being a solution, since it does not contain 5 nor 9.

Hence our aim is to get 2, 6, or 1.

However, 5 and 9 are coprime, and through the Euclidian algorithm,

2(5) − 9 = 1

Now that we have 1, we just need to add 5 twice, and get 11.

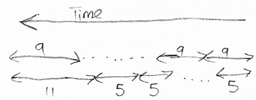
4(5) − 9 = 11

By the same reasoning as part 3, the solution is to start both timers together, and retiming the 5 minute timer when it stops. When the 9 min. timer stops, ~~play~~ start heating the water. The first time the 5 minute timer times out after start of heating, the water would have been heated 1 minute. Timing 5 minutes twice, we get 11 minutes.

Figure 6. William's solution - Stages III and IV

Alternative Solution 2

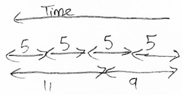
Since we are allowed to start the timer before the water is heated, we add 9 or 5 to 11, until it is a multiple of 5 or 9 respectively, this is because:



If we are able to do so, then we start timing whatever amount we added first before we start heating the water, as shown in the illustration.

By guess and checking,
Notice that $11+9=5(4)$

Hence we have the 9 minute played first before heating the water



Extension 1

Suppose we have 2 timers: one for x minutes, while the other for y minutes. For what values can I time?

By euclidean Algorithm,

$ax+by=\gcd(x,y)$, where a is the number of times x is used,
b " y " .

Hence the values I can time are multiples of the greatest common divisor of x and y.

Extension 2

Suppose I want to heat a beaker of water for t minutes, where t is prime. For what pairs of timer can I use?

From the previous extension, we know that for timers timing x and y minutes it must be a multiple of $\gcd(x,y)$.

Hence t must be a multiple of $\gcd(x,y)$.

But t is prime, hence $\gcd(x,y)=1$ or t.

Therefore it can only be timed by a pair of timers who's gcd is 1 or t.

Generalisation

There are 2 timers timing x and y minutes, and we want to heat up the water for z minutes. Assuming a solution exists, find the solution.

Let no. of times x is used be a.
" y " b.

$ax+by=1$.

$(az)x+(bz)y=z$

One of them is negative, and W.L.O.G, let bz be negative.

Then start both timers. When y is played bz times, start heating the water, with x still continuing. When x is played az times, the time z is used to heat the water, since $(az)x+(bz)y=z$.

Figure 7. William's solution

- alternative solutions, extension and generalization of the problem

William's solution was correct. In fact, it was a "good" solution that reflected advanced mathematical knowledge. William was in the school's mathematical Olympiad training team and had extra coaching in advanced mathematics. In Stage I (Understand the problem), William already demonstrated his attempt to attack the problem directly by identifying the attributes of the three given numbers: 5, 9 and 11 have no [non-trivial] common factors. In Stage II (Devise a plan) and Stage III (Carry out the plan), he presented very clearly his plan of tackling the problem using algebraic approach directly. Under Pólya's stage, he was awarded 10 marks (Level 3 of correct solution).

In Stage II (Devise a plan), he listed the heuristics/steps clearly (define variables; set up equations; solve the equation; related back to problem; offer solution). Furthermore, in Stage III (Carry out the plan), he formulated the Diophantine equation and solved it using the Euclidean Algorithm. This is evidence of rather sophisticated mathematical problem solving with rich resources (Schoenfeld, 1992), which is characteristic of an advanced mathematics student. Under Heuristics, he was awarded 4 marks (Level 2 of correct solution). Finally, in Stage IV (Check and Extend), William checked the reasonableness of his solution by a brief sensible argument. He was thus awarded 1 mark under Checking (Level 2). He offered two alternative solutions to this problem: one solution by considering the formation of 11 (shown under "Alternative Solution 1") and another solution by using a diagram. Under "Alternative Solutions" he was awarded 2 marks (Level 3). Even though William gave two extensions and one generalization under Check and Extend, these problems revolved around the same concept of the solution of the Diophantine equation $ax + by = d$. Under "Extending, Adapting & Generalizing", he was awarded 2 marks (Level 3). The detailed breakdown of his score is shown in Table 3.

Table 3

Score for William

	Descriptors	Marks awarded
Pólya's stages	Correct Solution – Level 3	10
Heuristics	Correct Solution – Level 2	4
Check and Extend	Checking – Level 2	1
	Alternative Solution – Level 2	2
	Extend, Adapt & Generalize – Level 2	2
	Total Score for William	19

8 Conclusion

While it is recognized that problem solving is "the heart of mathematics" (Halmos, 1980), and that there has been a worldwide push for problem solving to be the central focus of mathematics curriculum, implementation of problem solving has not met with much success. Schoenfeld (2007) puts matter-of-factly:

> That body of research – for details and summary, see Lester (1994) and Schoenfeld (1985, 1992) – was robust and has stood the test of time. It represented significant progress on issues of problem solving, but it also left some very important issues unresolved … The theory had been worked out; all that needed to be done was the (hard and unglamorous) work of following through in practical terms. (p. 539)

This chapter presents a conceptualization of mathematical problem solving based on a paradigm shift of learning and experiencing problem solving as part of a mathematics "practical". A corresponding scheme of assessment through the use of an assessment rubric is also introduced in this chapter.

The authors have carried out these problem solving practical lessons in a secondary school. As could be seen from the two students' scripts in the preceding section, the students were able to respond to the new mode of assessment as they were informed on how they would be assessed and shown the assessment rubric in a timely manner. Through the interviews reported in the previous section, the students were generally able to appreciate lessons on mathematical problem solving: the highest achieving student (represented by William) found that such practical lessons complement his mathematics Olympiad training received in school; the least interested student (represented by Zill) was able to appreciate the importance of heuristics in solving mathematics problems, even though he resented the problem solving lessons generally.

This new approach of assessing problem solving through the use of mathematics practical holds promise for teachers who want to elevate problem solving to a prominent position in the mathematics lessons. They can now not only encourage problem solving in their classes; they can also make transparent to students the criteria for assessment and the processes that are valued. As such, the practical worksheet can potentially become part of the overall assessment of students in their mathematics performance.

References

Carlson, M.P. & Bloom, I. (2005). The cyclic nature of problem solving: An emergent multidimensional problem-solving framework. *Educational Studies in Mathematics, 58*, 47-75.

Halmos, P. (1980). The heart of mathematics. *American Mathematical Monthly, 87*(7), 519-524.

Hedberg, J., Wong, K.Y., Ho, K.F., Lioe, L.T., & Tiong, Y.S.J. (2005). *Developing the repertoire of heuristics for mathematical problem solving: First Technical Report for Project CRP38/03 TSK.* Singapore: Centre for Research in Pedagogy and Practice, National Institute of Education, Nanyang Technological University.

Holton, D., Anderson, J., Thomas, B. & Fletcher, D. (1999). Mathematical problem solving in support of the classroom? *International Journal of Mathematical Education for Science and Technology*, *30*(3), 351- 371.

Lester, F.K. (1994). Musings about mathematical problem-solving research: 1970-1974. *Journal of Research in Mathematics Education, 25*(6), 660- 675.

Mason, J., Burton, L. & Stacey, K. (1985). *Thinking mathematically*. London: Addison-Wesley.

Ministry of Education. (2006). *A guide to teaching and learning of O-Level Mathematics 2007.* Singapore: Author.

Ministry of Education New Zealand. (2006). *Findings from the New Zealand Numeracy Development Projects 2005.* Wellington: Learning Media Limited.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Pólya, G. (1945). *How to solve it: A new aspect of mathematical method.* Princeton, NJ: Princeton University Press.

Schoenfeld, A. (1985). *Mathematical problem solving.* Orland, FL: Academic Press.

Schoenfeld, A.H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In: D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York: MacMillan Publishing.

Schoenfeld, A. (2007). The complexities of assessing teacher knowledge. *Measurement: Interdisciplinary Research and Perspectives, 5*(2/3), 198- 204.

Silver, E. A., Ghousseini, H., Gosen, D., Charalambous, C., & Strawhun, B. T. F. (2005). Moving from rhetoric to praxis: Issues faced by teachers in having students consider multiple solutions for problems in the mathematics classroom. *Journal of Mathematical Behavior*, *24*, 287-301.

Tay, E. G., Quek, K. S., Toh, L., Dong, F., & Ho, F. H. (2007). *Mathematical problem solving for integrated programme students: The practical paradigm*. Paper presented at the 4th East Asia Regional Conference on Mathematics Education, Penang, Malaysia.

Toh, T.L., Quek, K.S., & Tay, E.G. (2008). Mathematical problem solving – a new paradigm. In J. Vincent, R. Pierce, & J. Dowsey (Eds.), *Connected Maths* (pp. 356-365). Melbourne: Mathematical Association of Victoria.

Walvoord, B. E., & Anderson, V. J. (1998). *Effective grading: A tool for learning and assessment.* San Francisco, CA: Jossey-Bass Publishers.

Woolnough, B., & Allsop, T. (1985). *Practical work in science*. Cambridge: Cambridge University Press.

Appendix A

Practical Worksheet

Problem

Instructions

- You may proceed to complete the worksheet doing stages I – IV.
- If you wish, you have 15 minutes to solve the problem without explicitly using Polya's model. Do your work in the space for Stage III.
  - If you are stuck after 15 minutes, use Polya's model and complete all the stages I – IV.
  - If you can solve the problem, you must proceed to do stage IV – Check and Extend.

I Understand the problem

(You may have to return to this section a few times. Number each attempt to understand the problem accordingly as Attempt 1, Attempt 2, etc.)

(a) Write down your feelings about the problem. Does it bore you? scare you? challenge you?

(b) Write down the parts you do not understand now or that you misunderstood in your previous attempt.

(c) Write down your attempt to understand the problem; and state the heuristics you used.

Attempt 1

Page 1

II Devise a plan

(You may have to return to this section a few times. Number each new plan accordingly as Plan 1, Plan 2, etc.)

(a) Write down the key concepts that might be involved in solving the problem.

(b) Do you think you have the required resources to implement the plan?

(c) Write out each plan concisely and clearly.

Plan 1

Page 2

III Carry out the plan

(You may have to return to this section a few times. Number each implementation accordingly as Plan 1, Plan 2, etc., or even Plan 1.1, Plan 1.2, etc. if there are two or more attempts using Plan 1.)

(i) Write down in the *Control* column, the key points where you make a decision or observation, for e.g., go back to check, try something else, look for resources, or totally abandon the plan.

(ii) Write out each implementation in detail under the *Detailed Mathematical Steps* column.

Detailed Mathematical Steps	*Control*
Attempt 1	

IV Check and Expand

(a) Write down how you checked your solution.
(b) Write down your level of satisfaction with your solution. Write down a sketch of any alternative solution(s) that you can think of.
(c) Give one or two adaptations, extensions or generalisations of the problem. Explain succinctly whether your solution structure will work on them.

Appendix B

RUBRICS FOR ASSESSING PROBLEM SOLVING

Name: ______________________________

Polya's Stages		
	Descriptors/Criteria *(evidence suggested/indicated on practical sheet or observed by teacher)*	**Marks**
Correct Solution		
Level 3	Evidence of complete use of Polya's stages – UP + DP + CP; and when necessary, appropriate loops. [10 marks]	
Level 2	Evidence of trying to understand the problem and having a clear plan – UP + DP + CP. [9 marks]	
Level 1	No evidence of attempt to use Polya's stages. [8 marks]	
Partially Correct Solution *(solve significant part of the problem or lacking rigour)*		
Level 3	Evidence of complete use of Polya's stages – UP + DP + CP; and when necessary, appropriate loops. [8 marks]	
Level 2	Evidence of trying to understand the problem and having a clear plan – UP + DP + CP. [7 marks]	
Level 1	No evidence of attempt to use Polya's stages. [6 marks]	
Incorrect Solution		
Level 3	Evidence of complete use of Polya's stages – UP + DP + CP; and when necessary, appropriate loops. [6 marks]	
Level 2	Evidence of trying to understand the problem and having a clear plan – UP + DP + CP. [5 marks]	
Level 1	No evidence of attempt to use Polya's stages. [0 marks]	

Heurisitcs		
	Descriptors/Criteria *(evidence suggested/indicated on practical sheet or observed by teacher)*	**Marks**
Correct Solution		
Level 2	Evidence of appropriate use of heuristics. [4 marks]	
Level 1	No evidence of heuristics used. [3 marks]	
Partially Correct Solution *(solve significant part of the problem or lacking rigour)*		
Level 2	Evidence of appropriate use of heuristics. [3 marks]	
Level 1	No evidence of heuristics used. [2 marks]	
Incorrect Solution		
Level 2	Evidence of appropriate use of heuristics. [2 marks]	
Level 1	No evidence of heuristics used. [0 marks]	

Checking and Expanding		
	Descriptors/Criteria *(evidence suggested/indicated on practical sheet or observed by teacher)*	**Marks**
Checking		
Level 2	Checking done – mistakes identified and correction attempted by cycling back to UP, DP, or CP, until solution is reached. [1 mark]	
Level 1	No checking, or solution contains errors.[0 marks]	
Alternative Solutions		
Level 3	Two or more correct alternative solutions. [2 marks]	
Level 2	One correct alternative solution. [1 mark]	
Level 1	No alternative solution. [0 marks]	

Extending, Adapting & Generalizing		
Level 4	More than one related problem with suggestions of correct solution methods/strategies; or one **significant** related problem, with suggestion of correct solution method/strategy; or one **significant** related problem, with explanation why method of solution for original problem cannot be used. [3 marks]	
Level 3	One related problem with suggestion of correct solution method/strategy. [2 marks]	
Level 2	One related problem given but without suggestion of correct solution method/strategy. [1 mark]	
Level 1	None provided [0 marks]	

Hints given:

Marks deducted: ________________________

Total marks: ___________________________

Chapter 4

Assessing Conceptual Understanding in Mathematics with Concept Mapping

JIN Haiyue WONG Khoon Yoong

Mathematics educators and mathematics curriculum worldwide have emphasised the importance of students' ability to construct connections among mathematics concepts ("conceptual understanding") instead of just the competence to carry out standard procedures in isolated ways. Education researchers have used different techniques to assess this conceptual interconnectedness in students' minds. In this chapter, we discuss the use of concept mapping as an assessment tool in mathematics instruction, including different types of concept mapping tasks, training in concept mapping, applications in classroom settings, and evaluation of student-constructed concept maps. Concept mapping can be a worthwhile tool in teachers' repertoire of assessment for learning.

1 Introduction: What and Why of Concept Mapping

Cognitive psychologists have proposed that knowledge should be interconnected, and acquiring knowledge with understanding is to make meaningful connections between facts, concepts, and procedures. In mathematics, the importance of interconnectedness among mathematical concepts has been emphasized under the label "conceptual understanding" (Kilpatrick, Swafford and Findell, 2001; National Council of Teachers of Mathematics, 2000). For example, Van de Walle, Karp, and Bay-Willams (2010) define conceptual understanding as "the knowledge about relationships or foundational ideas of a topic" (p. 24), and these relationships are built from underlying concepts that are meaningful to the students. The Singapore mathematics syllabus (Ministry of Education, Singapore, 2006) highlights that students should develop a deep understanding of mathematical concepts and make sense of various mathematical ideas, including their connections and applications; that is, students should see mathematics as an integrated whole instead of isolated pieces of knowledge. Two principal issues to actualize this curriculum goal are finding ways to help students make connections among what they have learned and to assess their conceptual interconnectedness so that the information can be used by teachers to plan lessons and provide remediation. Education researchers have experimented with different techniques to assess conceptual interconnectedness (White and Gunstone, 1992), and this chapter addresses this assessment issue by focusing on concept mapping.

Concept mapping has gained popular use in science education over the past three decades, and is now being studied by mathematics educators (Afamasaga-Fuata'I, 2009). Figure 1 shows a concept map describing the relations among seven concepts related to triangles. As this figure shows, a concept map consists of three elements: (1) *nodes* representing concepts, usually enclosed in ovals or rectangles, (2) *links* showing connections between concepts, and (3) *linking phrases* specifying relationships between pairs of concepts. The nodes can be mathematical concepts, examples and non-examples of the concepts, diagrams, symbols, and formulas. The links are usually directional to show subject-object, pre-post, cause-effect, top-down hierarchy, or other relationships between the concepts. The linking phrases can be verbs or adjective phrases. When two or more nodes are linked, statements are formed, and these statements are called *propositions*. For example, in Figure 1, the connection between the concepts *triangle* and *acute-angled triangle* forms the proposition "*triangle*, when it has an acute angle, is an *acute-angled triangle*" (note that this proposition is only partially correct because *all* the angles of an acute-angled triangle must be acute). The propositions form the basic units of meaning in concept maps (Ruiz-Primo, 2004), although simpler concept maps may not have linking

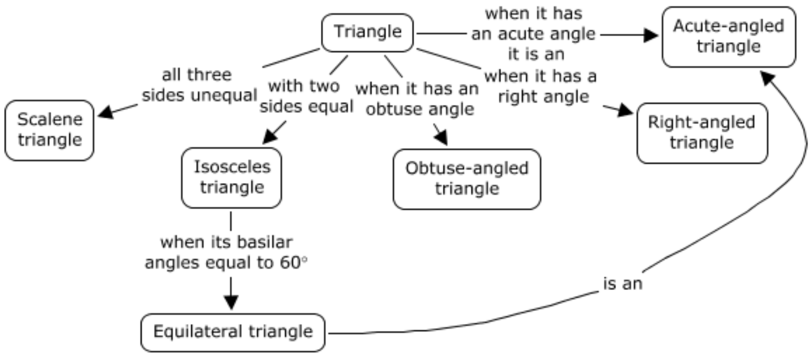


Figure 1. A concept map showing relations among concepts of triangles[1]

phrases (e.g., Orton, 2004), resulting in loss of information about the nature of the links.

Why is concept map useful in assessing conceptual understanding? Research suggests that the degree of a student's understanding is determined by the number, accuracy, and strength of connections (Hiebert and Carpenter, 1992; Resnick and Ford, 1981). Thus, a concept is well understood if it has sufficient number of accurate and strong links with other related concepts. From this perspective, a concept map can provide a visual representation of the interconnected properties of the concepts held by the student.

Concept map was first developed by Joseph Novak and his team in the 1970s as a tool to document changes in understanding of a wide range of scientific concepts held by students as they moved from first grade to twelfth grade (Novak and Musonda, 1991). It is based on Ausubel's (1963) *assimilation theory* that states that learning takes place by assimilating new concepts and propositions into existing knowledge framework or cognitive schema of the learner (Novak and Cañas, 2006). This psychological foundation justifies the use of concept map as a tool

[1] Most of the concept maps in this chapter are drawn using the software IHMC CmapTool, available at http://cmap.ihmc.us

to trace students' conceptual changes over time. Over the past three decades, its use as an assessment technique has been extensively investigated, especially in science education (Cañas, et al., 2003). In mathematics education, researchers and educators have also reported positive findings concerning the use of concept map as an assessment technique at different educational levels (see chapters in Afamasaga-Fuata'I, 2009). The following sections will discuss four aspects of this use.

First, several types of concept mapping tasks are explained to show that different tasks may address different aspects of students' conceptual understanding. Second, training techniques for the mapping tasks are illustrated with examples to help teachers plan such training when concept mapping is new to their students. Third, four different classroom applications of concept maps are discussed with examples, viz. to detect student's prior knowledge, to measure learning outcomes, to track learning progress, and to serve as a learning strategy. Finally, several methods to evaluate student-constructed concept maps are given so that the teachers can use the assessment information to plan meaningful learning; this will align with the perspective of assessment *for* learning.

2 Types of Concept Mapping Tasks

Concept mapping tasks can be categorized along a continuum of low to high degree of directedness according to whether the four components, namely, *concepts*, *links*, *linking phrases*, and *map structure*, are fully, partially, or not provided (Ruiz-Primo, Schultz, Li and Shavelson, 2001). In high-directed concept mapping tasks, most of these components are provided; thus, the tasks are relatively easy for students to complete, but they are limited in measuring the interconnected properties of students' understanding. In contrast, in low-directed concept mapping tasks, students have greater freedom to express their understating of a topic using components that they construct on their own. In this case, the openness of the tasks is more challenging to the students.

Some examples of commonly used concept mapping tasks from high-directed to low-directed ones are provided below.

2.1 *High-directed concept mapping tasks: Fill-in-the-map*

Fill-in-the-map tasks provide students with several concepts and require them to fill in a skeleton map with these concepts. Figure 2 shows two different examples of fill-in-the-map tasks: Fill-in-the-nodes and Fill-in-the-lines. Distracters may be included to encourage students to think carefully about which items are relevant to the map. The Fill-in-the-nodes task in Figure 2 is an incomplete concept map with two blank nodes. Four concepts are provided, two of which are distracters. On the other hand, the Fill-in-the-lines task has two unlabelled links. Two linking phrases are provided, with no distracter. The teacher has to decide whether or not to include distracters, depending on the stage of learning and the students' ability. In either case, students fill in the blanks with what they think are the correct items based on their understanding.

To design this type of mapping task, teachers either construct a concept map themselves or use an expert-constructed map (for example, through working with other teachers or mathematicians). Then remove some of the concepts or linking phrases from the map and add distracters, if desired. This type of concept mapping task is easy to administer and to grade, for example, by counting the number of correctly filled items.

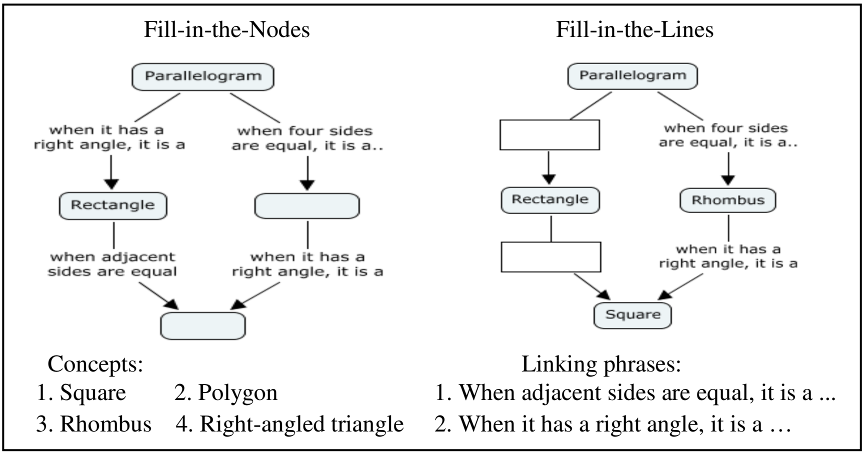


Figure 2. Examples of fill-in-the-nodes and fill-in-the-lines skeleton maps

2.2 *Semi-directed concept mapping tasks*

When one or two of the above mentioned four components of a complete concept map is missing and the other remaining components are fully or partially provided, the concept mapping task is considered to be *semi-directed*. Compared with the high-directed concept mapping tasks, the semi-directed mapping tasks require more efforts to complete.

In the semi-directed concept mapping task shown in Figure 3, only concepts and linking phrases are provided. Students need to construct a concept map including all the given concepts but only the relevant linking phrases. An example of a possible concept map is also shown.

Please construct a concept map using the concepts in the list given below. A list of linking phrases is provided to help you find the possible connections between the concepts. However, you do not need to use all the linking phrases for your concept mapping. You are also allowed to add extra but related concepts and use your own linking phrases if no appropriate one is found from the given linking phrases list.

Concepts:
composite numbers
whole numbers
prime numbers
zero
factor

Linking phrases:
… can be divided into
… include …
… does not include …
… is not a …
… is a ….
… have more than two different …
… have only two, 1 and itself …
… belongs to …

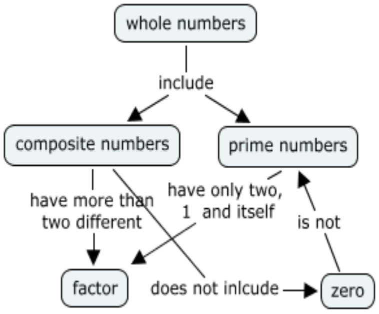


Figure 3. An example of semi-directed concept mapping task

A variation of a semi-directed concept mapping task, illustrated in Mansfield and Happs (1991), is to provide students with a partial list of concepts of a particular topic, say, quadrilaterals. The most inclusive concept, in this case, *quadrilateral,* is placed at the top of the map, with the other less inclusive concepts, such as *rectangle* and *rhombus,* at lower levels, thereby requiring students to consider hierarchy among the concepts. Empty boxes are provided for students to fill in the concepts and linking phrases.

2.3 *Low-directed concept mapping tasks: Free-style mapping*

Low-directed concept mapping tasks, also called free-style mapping, require students to fully construct the maps based on the mapping topic or a list of given concepts. They are free to express ideas in their own ways covering the four components of a concept map. When only a mapping topic is given, students need to first identify some concepts relevant to the topic and then construct the map accordingly. For most school students, a concept list is usually given because they may have difficulty in selecting the appropriate concepts. Some of them may provide concepts that are somewhat related but not relevant or essential to the topic (Jin, 2007). For example, they may include the concept *mathematics* within the topic of *functions*: "functions are very important for the learning of mathematics". This kind of propositions does not directly address students' understanding about *functions,* and irrelevant concepts may even distract students from constructing meaningful maps. A given concept list will help them focus on a certain knowledge domain; at the same time, the task can allow students to include additional concepts that they think are relevant to the given ones.

The concept map in Figure 4 by a Singapore Secondary 3 student is an example of a low-directed concept mapping task with a given list of concepts about quadrilaterals. The student had used all the concepts provided without adding new ones. The map was well constructed, with the most general concept *polygon* located at the top, followed by less inclusive concepts *quadrilateral, parallelogram,* and similar shapes at the middle levels, and the least inclusive concept *diagonals* at the bottom.

Concept mapping task: Construct a concept map using the given concepts: quadrilateral, polygon, kite, parallelogram, rectangle, square, diagonals, rhombus, trapezium

(A student-constructed map example is shown below – redrawn for clarity)

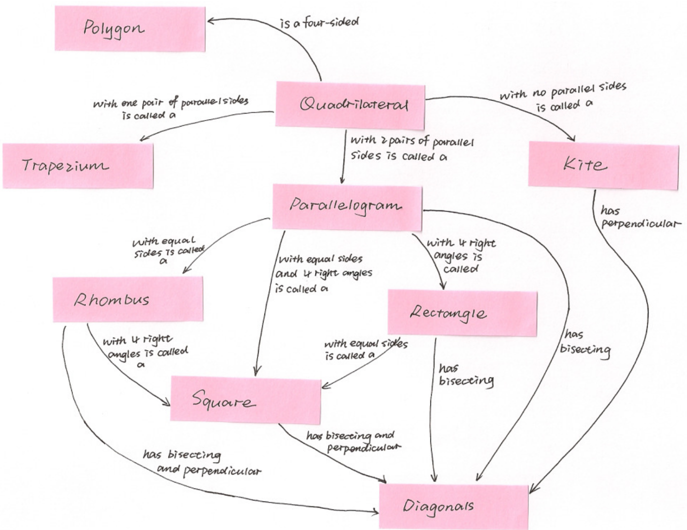


Figure 4. An example of low-directed concept mapping task about quadrilaterals

Teachers may begin with high-directed mapping tasks and then move to low-directed ones. This is because the high-directed mapping tasks are relatively easy for students to complete. Furthermore, starting with easier tasks allows time for both teachers and students to become familiar with the purposes and construction of concept maps before they tackle the more challenging low-directed tasks.

3 Training on Concept Mapping

Where concept map has not been extensively used in mathematics lessons, it is necessary to train students on the techniques of constructing informative concept maps. High-directed and semi-directed mapping tasks are, however, quite straightforward and do not require extensive training. Thus, this section will focus on training in free-style mapping tasks where a list of concepts is given. This is also most commonly used by researchers and widely reported in the literature.

The following procedures have been developed based on the literature and pilot studies conducted in Singapore and China (Jin and Wong, 2010).

1. **Introduction**: Teachers first provide students with a preliminary idea of what a concept map is, what it is used for, and what its attributes are, i.e., nodes, arrowed links, and linking phrases.
2. **Demonstrate with examples**: Teachers begin with an example with four or five concepts that students have already learned. First, read aloud the concepts and help students to recall their meanings. Second, write the concepts onto separate cards so that they can be easily moved around to explore various connections that make sense. Concepts that are most closely related are arranged near to one another. Third, once the intended connections have been decided upon, identify the relations between each pair of concepts and draw directed lines between them. Fourth, write on each line the relationships identified so that propositions can be formed. Finally, go back and check to see if any concept or relationship has been left out; redraw the map if necessary.
3. **Student practice**: Provide students with a different set of concepts for practice and remind them to pay attention to the following:
    a. All the given concepts should be included in the map.
    b. In arranging the concepts, make sure enough space is left for adding linking phrases.
    c. The lines should be directed (with arrow) so that the relationships are clear.
    d. All the lines should be labeled with linking phrases.
    e. The entire map should be clear and legible.
4. **Consolidation**: After some practice, students should have mastered the basic skills of concept mapping. Teachers should

further encourage students to include additional relevant concepts into their concept maps, to construct as many relationships as they can between the concepts, and to describe the relationships using informative, detailed linking phrases. As the concept map is used as a graphical representation of students' conceptual understanding, they should show as much of their knowledge as possible in their map so that it is rich enough to capture the essential attributes of their conceptual understanding. With this the teachers can obtain a better idea about what the students have already grasped and which contents they are still weak in.

Some researchers emphasize the hierarchical nature of concept maps because of Ausubel's learning theory that more general, superordinate concepts should subsume more specific, detailed concepts. In mathematics, in light of Skemp's (1986) *Schema Theory*, a concept map that shows the hierarchy of the concepts is a more comprehensive representation of the interrelatedness among mathematical concepts. However, strict requirement on hierarchy may distract students from constructing meaningful connections, which is the main concern of most mapping tasks at the school level. Besides, some school students might have difficulty distinguishing or expressing the hierarchy of abstract mathematical concepts (Schau and Mattern, 1997). Thus, for primary or secondary school students, it is appropriate to encourage rather than require them to construct concept maps with strong hierarchy.

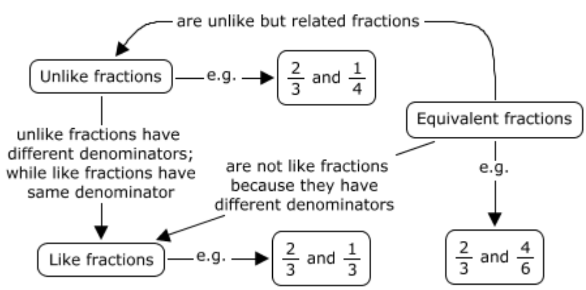
4 Classroom Applications of Concept Map

This section covers four different but related applications of concept map in classroom teaching and assessment.

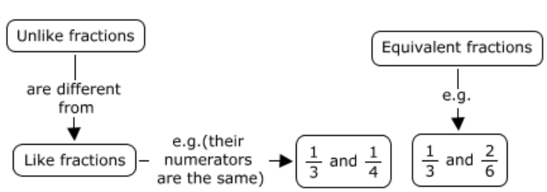
4.1 *Using concept map to detect students' prior knowledge*

The prior knowledge that students bring to their learning experience affects how they encode and later retrieve newly learned information (Dochy, 1994). Concept map has been used to find out about this prior knowledge (DiCerbo, 2007; Gurlitt and Renkl, 2008) so that more effective lessons and materials can be prepared to link prior knowledge to new learning.

For example, before teaching "addition of two unlike fractions", teachers need to know what their students have mastered about prior concepts such as *like fractions, unlike fractions, equivalent fractions*, and so on. They may design one type of mapping task to do so. Take the case of a low-directed concept mapping task. The teacher can ask students to construct concept maps using a list of concepts as given or selected by the students. Figure 5 shows two student-constructed concept maps with the above three concepts about fractions. Student A has displayed clearly the correct relationships among the three concepts and included relevant numerical examples. By contrast, Student B has not mentioned substantial relationships among the concepts; the only relationship "*unlike fractions* are different from *like fractions*" is very brief and general. Furthermore, the example for *like fractions* is wrong, equating the numerators rather than the denominators. Thus, Student B will have difficulty learning the new topic on unlike fractions, and some remediation is necessary.



Student A's map



Student B's map

Figure 5. Two examples of student-constructed concept maps with given concepts

4.2 *Using concept map to evaluate learning outcomes*

Concept mapping tasks can be assigned to students to assess their understanding of newly-taught concepts. To avoid compounding conceptual mapping with the new learning, it is better to begin with semi-directed concept mapping tasks. Figure 6 is an example measuring students' understanding after learning *numbers*, modelled after Mansfield and Happs (1991). In this task, some related concepts are given in boxes while others are omitted, with the boxes left blank for students to fill in. The omitted concepts are provided on the right-hand side, together with some distracters. The hierarchical positions of the concepts are fixed as given. To some extent, these positions give hints to the appropriate concepts for the blank boxes. For example, those who know the relationship between *composite number* and *prime number* can deduce that the blank box next to *composite numbers* should be *prime numbers*. In addition to the blank boxes, spaces are provided for students to add linking lines and labels in their own words.

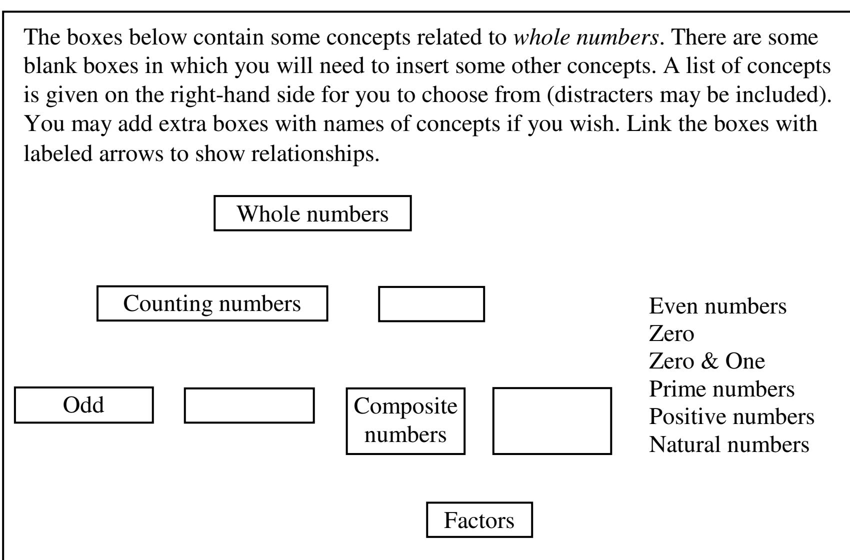


Figure 6. A semi-directed concept mapping task on whole numbers

With the given concepts, either in the boxes or in the list, and the fixed hierarchical positions of the concepts, this task focuses students' attention on a particular domain. Thus, teachers have more control over what they are testing. This task can be converted to the low-directed type, for example, by providing only the mapping topic *numbers* or offering a list of concepts related to *numbers*. As mentioned in the earlier section, this new task is more challenging for students to complete and teachers to grade; yet, its openness allows students to have greater freedom to express their understanding, thus, giving more valuable information to the teachers.

4.3 *Using concept map to track students' progress in learning*

Concept map has been used to track the changes in a student's cognitive structure and its increasing complexity as he or she integrates new knowledge into existing cognitive structure. For example, Mansfield and Happs (1991) reported their use of concept map as an expedient evaluation tool in a study of parallel lines among a group of 12-year old students. In their study, the same concept list consisting of eleven concepts was provided in the pre- and post- concept mapping tests before and after instruction of parallel lines. They cited the concept maps drawn by the student, Bruce: he used only five concepts in the pre-concept map but seven, adding one new concept, in the post-concept map. Furthermore, the propositions in the post-concept map were more informative, for example, revealing a misconception not found in the pre-concept map. Although the two maps were well-constructed, the omission of some given concepts suggests that further teaching of these concepts is required for that student. Thus, comparing concept maps constructed before and after instruction can help teachers determine how much progress their students have made and how effective the instruction has been. With the information drawn from the comparison, teachers can then adjust their plans for future lessons.

4.4 *Constructing concept maps as a learning strategy*

Concept mapping has been widely promoted as a learning strategy to help students elaborate on their learning and thereby to develop deep conceptual understanding (Afamasaga-Fuata'I, 2006; Jegede, Alaiyemola and Okebukola, 1990). This popular application extends concept mapping beyond its use as a formative or summative assessment tool.

Concept map can serve as a scaffold for students to organize their own knowledge. At the end of a unit of instruction, a high-directed concept mapping task can help students clarify the nature and number of connections among newly learned concepts. A low-directed concept mapping task will encourage them to reflect on the possible relationships among the concepts and represent these relationships in a pictorial way. They may even see links that they are not initially aware of (De Simone, 2007), thereby developing deeper understanding about the concepts. They can modify these maps as learning progresses. This constructive activity can be more effective when students work together in groups to discuss their ideas and combine their knowledge in order to learn and construct new knowledge (Gao, Shen, Losh and Turner, 2007). This group activity will also provide opportunity for groups to compare their conceptual structure with other groups and this will inspire further learning. In recent years, with the support of concept mapping software such as CmapTools (Novak, 1998) and SmartDraw (http://www.smartdraw.com), students can now build and discuss their concept maps at distant locations (Novak and Cañas, 2006) and flexible times.

Several studies (e.g., Kankkunen, 2001; Mohamed, 1993) have reported students' positive attitudes toward using concept map in science. In mathematics, we conducted a study with a class of Grade 8 Chinese students (n = 48) in 2009. The students' attitudes toward concept map were collected through a questionnaire and interviews after their one month's experience with concept mapping. Most of them agreed that concept mapping was useful in learning mathematics. They expressed moderate to high levels of enjoyment of concept mapping even though, at the same time, some of them admitted that concept mapping was challenging and required hard thinking. These findings are encouraging for teachers who wish to explore this technique in their mathematics lessons.

5 Evaluation of Student-Constructed Concept Maps

Concept maps can be assessed in a holistic, qualitative way based on expert or teacher impressions or scored using specific criteria. These methods should result in meaningful grades or scores so that judgement about the quality of students' conceptual understanding can be made from concept maps.

The following sections describe several quantitative methods to score concept maps by examining the links between individual concepts and the quality of the whole map. These scores can be used to assess students' performance on the given concept mapping task. It is not necessary to use all the methods below for classroom assessment; however, some of these methods may be used in action research.

5.1 *Links between concepts*

As defined in the introduction section, a concept map is a directed network. The number of links, including incoming and outgoing ones, connected to an individual concept reflects the extent to which that concept connects to all the other concepts in the network, with higher number of links showing that it has stronger connections with other concepts in the domain. This will reflect the students' conceptual understanding of that concept. In the extreme case, an isolated concept with no incoming and outgoing link suggests that the person is not familiar with the concept, cannot recall the link, or has simply forgotten to construct connections with it (which can happen under timed test conditions). For missing links, the teachers may need to interview the students to find out the reasons behind their lack of conceptual connections about the concepts.

Tables 1 and 2 show the number of links to the concepts (excluding examples) found in the respective maps in Figure 5. Both maps have included all the three given concepts. Each concept in Student A's map has more links compared to the same concept in Student B's map. If the links are also mathematically correct, then Student A has a better understanding of the fraction concepts than Student B.

Table 1
Number of links to concepts in Student A's map in Figure 5

Concepts	Incoming links	Outgoing links	Total links
Like fractions	2	0	2
Unlike fractions	1	1	2
Equivalent fractions	0	2	2

Table 2
Number of links to concepts in Student B's map in Figure 5

Concepts	Incoming links	Outgoing links	Total links
Like fractions	1	0	1
Unlike fractions	0	1	1
Equivalent fractions	0	0	0

In addition to the number of links for each concept, it is also informative to examine the connections between pairs of concepts, in order to find out, for examples, how close or far apart are the two concepts (called the *distance*), how strong is the connection in terms of the number of direct and indirect paths between them (called the *connectedness*), and the quality of the *propositions.* The *distance* and *connectedness* measures indicate how easy or difficult it is for the students to access the respective pair of concepts in their cognitive structure. Knowing which pairs of concepts are cognitively "far apart" or "weak" in their students' conceptual understanding will alert the teachers to plan more focussed activities that can strengthen these particular links. These quantitative measures are related to various indicators used in social network analysis (Degenne and Forsé, 1999).

The *distance* between two concepts i and j is defined based on the length of the shortest path that connects them in the direction from i to j. When there is a direct link from i to j without any in-between concept, their distance $d(i, j) = 1$; when i and j are not connected, either directly or indirectly, their distance is defined as zero. Hence, it is possible to have $d(i, j) = 1$ and $d(j, i) = 0$. When there is an indirect link from i to j, their distance equals to the number of in-between concepts in the shortest path plus one. For any two connected concepts, the larger their distance, the further apart they are, and the harder for the students to connect them in their thinking. For the two maps in Figure 5, the *distance* matrices are given in Tables 3 and 4 respectively. For these simple maps, all the connections between pairs of concepts are direct ones ($d = 1$). It is also evident that Student A has constructed more links than Student B.

Table 3

Distance of the concepts in Student A's map in Figure 5

From / To	Like fractions	Unlike fractions	Equivalent fractions
Like fractions		1	1
Unlike fractions	0		1
Equivalent fractions	0	0	

Table 4

Distance of the concepts in Student B's map in Figure 5

From / To	Like fractions	Unlike fractions	Equivalent fractions
Like fractions		1	0
Unlike fractions	0		0
Equivalent fractions	0	0	

The *connectedness* between any two concepts counts the number of different paths between them. Referring to Student A's map in Figure 5, the *connectedness* from *equivalent fractions* to *like fractions* is 2 since there are two paths, one direct (*equivalent fractions* → *like fractions*) and one indirect (*equivalent fractions* → *unlike fractions* → *like fractions*); the *connectedness* of the same pair of concepts in Student B's map is 0, since there is no path between them. Pairs of concepts with large *connectedness* are quite robust; when one connection is broken, these concepts still have high chances of being linked together. Tables 5 and Table 6 show the corresponding *connectedness* matrices for the maps in Figure 5. The pairs of concepts in Student A's map have stronger or more robust relationships than those in Student B's map.

Table 5

Connectedness of concepts in Student A's map in Figure 5

From / To	Like fractions	Unlike fractions	Equivalent fractions
Like fractions		1	2
Unlike fractions	0		1
Equivalent fractions	0	0	

Table 6

Connectedness of concepts in Student B's map in Figure 5

From / To	Like fractions	Unlike fractions	Equivalent fractions
Like fractions		1	0
Unlike fractions	0		0
Equivalent fractions	0	0	

The third measure of the relationship between each pair of concepts is the quality of the *proposition* indicated by the linking phrase between them. A simplified scoring scheme is as follows (see McClure, Sonak and Suen, 1999; Novak and Gowin, 1984):

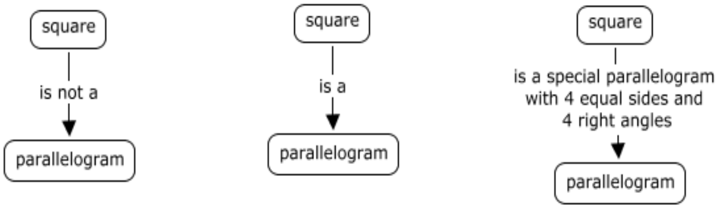
(1) when a proposition has no linking phrase or indicates misconception, score 0;

(2) when a proposition indicates a relationship between the

connected concepts but with partially correct or incomplete linking phrases, score 1;

(3) when a proposition indicates a correct and meaningful relationship between the connected concepts, score 2.

This scheme is illustrated in Figure 7 below. Proposition (*a*) indicates a misconception of the relationship between *square* and *parallelogram;* thus, it is scored 0. Proposition (*b*) correctly indicates the relationship that "*square* is a *parallelogram*"; however, the linking phrase is very brief and gives no further information about how or why a *square* is a *parallelogram.* Compared to proposition (*b*), the proposition (*c*) is more detailed as it gives a complete description of a *square* in relation to *parallelogram.* To apply the scoring scheme, teachers should decide how much their students are expected to master the relationship between two concepts at their stage of learning. If the students' propositions have met the expectation, then score 2; if the expectation is only partially met, score 1. With this in mind, a teacher may score proposition (*b*) 1 or 2 according to his/her expectation. For formative assessment, pay attention to the actual propositions, in addition to the scores, in order to identify students' good understanding as well as misconceptions.



(*a*) (*b*) (*c*)

Figure 7. Examples for scoring of propositions

The usefulness of the *distance*, *connectedness,* and quality of *proposition* scores becomes apparent for more complex concept maps. These scores allow more objective comparison between student-constructed concept maps.

5.2 *Nature of the whole map*

Student-constructed concept maps can be compared at the whole map level. The notion of *density* and the sum of all separate proposition scores can be used to indicate the holistic properties of a concept map.

The *density* refers to the ratio of the total number of links to the total number of concepts in a concept map. This provides information about how compact the concepts are tied together within a particular group in the map. There is no expected value of what the *density* of a concept map could reasonably be, but compact maps are likely to have strongly intertwined associations in the person's cognitive structure. For example, the densities of the concept maps in Figure 5 are $\frac{3}{3} = 1$ and $\frac{1}{3}$ respectively, excluding examples and the links to examples. This suggests that Student A's map is more compact than Student B's. Nevertheless, a higher density does not necessarily indicate better quality of a map since students may simply draw links without considering whether the links are substantial (meaningful) or trivial. The scoring of propositions helps to show this differentiation. Hence, the sum of all separate proposition scores is the second measure of the quality of the whole map. For example, in Figure 5, Student A's map obtained $2 + 2 + 2 = 6$ points as an overall proposition score since all the three propositions in the map are substantial ones; while Student B's map obtained 1 point since there was only one partially correct proposition between the three given concepts. In general, high proposition sums are associated with competent students who can provide many valid content-based propositions, whereas low sums are associated with weak students who do not provide many meaningful propositions.

Meaningful comparisons between concept maps can be made only if they cover the same given concepts. This is because some concepts are less compact than other concepts. As a consequence, a concept map constructed with such concepts will have fewer expected connections and therefore lower density and proposition scores. A different approach is to compare student-constructed concept maps against a criterion (or expert) map. The criterion map can be constructed by one or more teachers by taking into consideration the learning objectives. Any gaps between student maps and the criterion map (for examples, isolated concepts in student maps) and student misconceptions will highlight where further teaching is to be focussed on. Some students may achieve higher scores than the criterion map if they have constructed “insightful” connections that the teachers have not thought about; indeed, this shows that teachers may learn from their students.

6 Conclusions

This chapter has described three different types of concept mapping tasks that can be used as alternative assessment to supplement traditional paper-and-pencil tests, with concept maps highlighting the degree of conceptual understanding while traditional tests covering standard skills and problem solving. Of these three types, student-constructed concept map is particularly suited to measure individual conceptual understanding of a set of concepts. By asking students to explicitly consider how and why concepts are linked, teachers can detect students’ progress and gaps in understanding and then adjust their instruction accordingly. At the same time, the concept mapping tasks provide students important learning opportunity to reflect on what they have learned and help them see links that they may have missed.

Training students to construct their own concept maps and interpreting these maps are likely to be time-consuming. Some efforts may be saved by using simple scoring as explained above. Even so, it is not “economical” in terms of curriculum time to use concept maps for assessment purposes only. The studies on concept maps in mathematics (Afamasaga-Fuata’I, 2009; Schau and Mattern, 1997) have provided strong evidence of the advantages of using concept mapping as a teaching and learning tool, including letting students construct group concept maps. Once concept map has been used for instruction or learning, the burden of training for assessment will be reduced as students will have become familiar with features of concept maps that will be assessed. Hopefully, the students will also be receptive to the idea of using concept mapping as part of standard assessment. This will change concept mapping from assessment *of* learning to assessment *as* learning.

In summary, concept mapping is an assessment technique that can be applied at various stages of learning. The increasing number of studies in recent years on the uses of concept map in mathematics and other subjects suggests that students can benefit from concept mapping. Its effects on learning are also well-documented. Thus, it is worthwhile for teachers to develop their skills in using this assessment technique and to explore its use in their mathematics lessons to meet the curricular goals of promoting conceptual understanding in mathematics.

References

Afamasaga-Fuata'I, K. (2006). Developing a more conceptual understanding of matrices & systems of linear equations through concept mapping and Vee diagrams. *Focus on Learning Problems in Mathematics, 28*(3&4), 58-89.

Afamasaga-Fuata'I, K. (Ed.). (2009). *Concept mapping in mathematics: Research into practice.* New York: Springer.

Ausubel, D. (1963). *The psychology of meaningful verbal learning*. New York: Grune & Stratton.

Cañas, A.J., Coffey, J.W., Carnot, M.J., Feltovich, P., Hoffman, R.R., Fletovich, J., & Novak, J.D. (2003). *A summary of literature pertaining to the use of concept mapping techniques and technologies for education and performance support* (Report to The Chief of Naval Education and Training). Pensacola, FL: Institute of Human and Machine Cognition.

Degenne, A., & Forsé. M. (1999). *Introducing social networks* (A. Borges, trans). London: SAGE Publications. (Original work published 1994)

De Simone, C. (2007). Applications of concept mapping. *College Teaching, 55*(1), 33-36.

DiCerbo, K. E. (2007). Knowledge structures of entering computer networking students and their instructors. *Journal of Information Technology Education*, *6*, 263–277.

Dochy, F. (1994). Prior knowledge and learning. In T. Husen & T.N. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed.) (pp. 4698-4702). Oxford/New York: Pergamon.

Gao, H., Shen, E., Losh, S., & Turner, J. (2007). A review of studies on collaborative concept mapping: What have we learned about the technique and what is next? *Journal of Interactive Learning Research, 18*(4), 479-492.

Gurlitt, J. & Renkl, A. (2008). Are high-coherent concept maps better for prior knowledge activation? Different effects of concept mapping tasks on high school vs. university students. *Journal of Computer Assisted Learning, 24,* 407-419.

Hiebert, J., & Carpenter, T. (1992). Learning and teaching with understanding. In D. Grouws (Ed.), *Handbook of research in mathematics teaching and learning* (pp. 65-100). New York: Macmillan.

Jegede, O. J., Alaiyemola, F. F., & Okebukola, P. A. (1990). The effect of concept mapping on students' anxiety and achievement in biology. *Journal of Research in Science Teaching, 27,* 951-960.

Jin, H. (2007). *On the internal networks of middle school students' mathematics knowledge: Elementary function* [In Chinese 中学生函数知识网络的调查研究]. Unpublished master's thesis, Nanjing Normal University, Nanjing, China.

Jin, H., & Wong, K. Y. (2010). Training on concept mapping skills in geometry, *Journal of Mathematics Education, 3*(1), 103-118.

Kankkunen, M. (2001). Concept mapping and Peirce's semiotic paradigm meet in the classroom environment. *Learning Environments Research, 4,* 287-324.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.) (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academic Press.

Mansfield, H., & Happs, J. (1991). Concept maps. *The Australian Mathematics Teacher, 47*(3), 30-33.

McClure, J.R., Sonak, B., & Suen, H.K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching, 36*(4), 475-492.

Ministry of Education, Singapore. (2006). *Mathematics syllabus: Primary*. Singapore: Author.

Mohamed, N.B.R.A. (1993). *Concept mapping and achievement in secondary science*. Unpublished master's thesis. Singapore: National University of Singapore.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Novak, J. D. (1998). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations.* Mahwah, NJ: Lawrence Erlbaum.

Novak, J. D., & Cañas, A. J. (2006). The theory underlying concept maps and how to construct them [Electronic Version]. *Technical Report IHMC CmapTools 2006-01, Florida Institute for Human and Machine Cognition*. Retrieved December 1, 2008, from http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlying Concept Maps.pdf

Novak, J.D., & Gowin, D. B. (1984). *Learning how to learn.* London: Cambridge University Press.

Novak, J.D., & Musonda, D. (1991). A twelve-year longitudinal study of science concept learning. *American Educational Research Journal, 2*, 117-153.

Orton, A. (2004). *Learning mathematics: Issues, theory and classroom practice* (3rd ed.). London: Continuum.

Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Erlbaum.

Ruiz-Primo, M.A. (2004). Examining concept maps as assessment tool. In A.J. Cañas, J.D. Novak & F.M. Gonzalez (Eds.), *Concept maps: Theory, methodology, technology.* Proceedings of the First Conference on Concept Mapping. Retrieved March 25, 2008, from http://cmc.ihmc.us/CMC2004Programa.html

Ruiz-Primo, M.A., Schultz, S.E., Li, M., & Shavelson, R.J. (2001). Comparison of the reliability and validity of scoring from two concept-mapping techniques. *Journal of Research in Science Teaching, 3*(2), 260-278.

Schau, C., & Mattern, N. (1997). Use of map techniques in teaching statistics courses. *The American Statistician, 51*(2), 171-175.

Skemp, R. R. (1986). *The psychology of learning mathematics* (2nd ed.). Middlesex, England: Penguin Books.

Van de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2010). *Elementary and middle school mathematics: Teaching developmentally* (7th ed.). Boston, MA: Allyn & Bacon.

White, R., & Gunstone, R. (1992). *Probing understanding*. London: Falmer Press.

Chapter 5

Using Journal Writing to Empower Learning

Berinderjeet KAUR CHAN Chun Ming Eric

Amongst the number of alternative assessment methods that mathematics teachers can use, journal writing apparently can be said to be the easiest assessment method to carry out in the mathematics classroom without having to compromise significantly the teachers' formal teaching time. Because of its flexibility and ease of use and the benefits that can be reaped towards achieving learning outcomes, journal writing should be seriously considered as having a place in the mathematics classroom. This chapter illustrates with the use of students' sample responses to journal prompts how students may be engaged in writing about their understanding of what they are taught, mathematical content, processes, application, and attitude. In addition, it looks at how journal writing may be evaluated so that constructive feedback is given to the students. Finally, some pitfalls to avoid when implementing journal writing are discussed.

1 Introduction

Assessment is an integral part of teaching and learning. Essentially, assessment is a means of gathering information about students' progress with respect to achieving learning goals and providing feedback to inform instruction. Assessment has gone beyond the traditional sense of "testing" where a test score or grade is seen as final. Today, it is viewed as a dynamic process that includes a range of assessment strategies that will provide a repertoire of evidence to suggest students' learning and growth on a timely basis (National Council of Teachers of Mathematics, 1995; Ministry of Education, 2000a and 2000b). With emphasis on mathematical thinking and communication, there is a need to describe, explain, argue, and interpret in reformed classrooms. In such classrooms, alternative forms of assessment are seen as providing a more complete and accurate picture of students' learning and performances. The Ministry of Education (MOE) in Singapore has embraced this shift in focus towards a holistic perspective to assessment that includes assessment modes such as mathematical investigation, journal writing, classroom observation, self-assessment and portfolio assessment (MOE, 2004a and 2004b). This chapter focuses on journal writing as an alternative assessment strategy.

2 Review of Literature

Journal writing involves students writing about their mathematics learning. It is a systematic way of documenting learning and collecting information for self-analysis and reflection (Kaur, Ang, Kang, Ng, Ng, Phua, et al., 2004; Kaur, Chan, and Edge, 2006). According to the Assessment Guides to Primary Mathematics and Lower Secondary Mathematics, it is a platform where "students write about a mathematics-related situation in a way that reveals their learning of certain mathematics concepts, how it was taught, the difficulties encountered, and their feelings and attitudes towards the learning" (MOE, 2004a, p. 57 and 2004b, p. 54). In the Principles and Standards for School Mathematics of the US it is emphasised that

> reflection and communication are intertwined processes in mathematics learning … Writing in mathematics can also help students consolidate their thinking because it requires them to reflect on their work and clarify their thoughts about the ideas. (National Council of Teachers of Mathematics, 2000, p. 61)

According to Ernig (1977), writing provides information processing at the motor level (hand moving the pen), the sensory level (eyes reading), and the cognitive level (intellectual and analytical processing of the message).

Studies have documented positive effects of journal writing in mathematics classrooms with respect to achievement (Evans, 1984; Shepard, 1993), mathematical reasoning (Pugalee, 2005) and problem solving (Bell and Bell, 1985; Pugalee, 2005). Borasi and Rose (1989) asserted that journal writing in mathematics teaching had beneficial therapeutic effect on the feelings and attitudes of the students, as well as positive effect on their learning of mathematical concepts and problem solving skills. They also asserted that the interaction of students and teachers through journal writing may produce a beneficial supportive class atmosphere. Burns (2004) in her work with students on writing in mathematics noted that

> Not only did I see how writing helped students think more deeply and clearly about mathematics, but I also discovered that students' writing was an invaluable tool to help me assess their learning. (p. 30)

Studies in Singapore on the use of journal writing in mathematics lessons too revealed positive outcomes such as an effective way to increase teacher's understanding of students' learning in mathematics, attitudes and dispositions towards mathematics (Chai, 2004) and higher gains in mathematics test results (Ng, 2004; Yeo, 2008).

Clarke, Waywood, and Stephens (1993) studied the use of journal writing in mathematics to foster the development of metacognitive processes. A major finding of their long-term journal writing study was that students convincingly explained why they used journal writing:

> Sixty percent of the students gave as the main reason for writing in their journal, *because it helps me* (…), the most popular justification for journal use was *To help me learn* (…). Half of the student sample reported that the most important thing learned from journal completion was *To be able to explain what I think.* (p. 241)

In a three year assessment project (2005-2008) led by Professor Magdalena Mok (Mok, 2010) in Hong Kong, parents too realised the value of journal writing by their children. One parent stated that "after two years, I gradually realise the benefits of the Mathematics Journal" (Mok, 2010, p. 73). The parent also stated the following benefits:

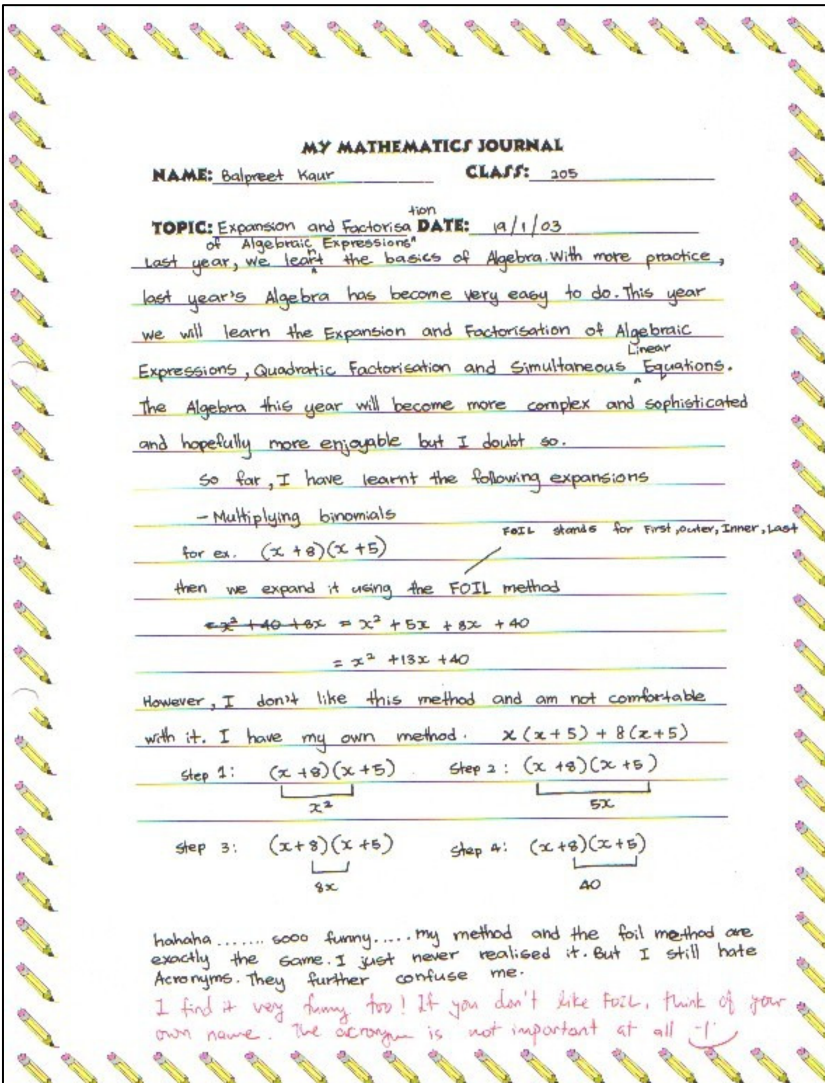
- Mathematics Journal is actually a cognitive training: students can reflect on the principles and methods of computation while they write the Mathematics Journal, such that mathematics understanding is consolidated.
- Mathematics Journal can help early reflection on one's mistakes: Very often, students would think that if they get the answer correct then they know how to do mathematics. In reality, they are discouraged or give up when they meet complicated items requiring comprehension. Only through the Mathematics Journal do they know how to reflect on their mistakes and put deep thoughts to the items. (pp.73-74)

3 Two Types of Journal Writing in the Mathematics Classroom

Selecting an appropriate assessment method depends on the purpose of the assessment. The brief review of literature in the previous section has provided a sense of what journal writing can be intended for. This section exemplifies free writing and writing from a prompt as two possible ways to engage students in journal writing.

3.1 *Free writing*

Students may be asked to write journals at regular intervals to keep track of their thoughts and experiences during mathematics lessons. The teachers may not provide them with any specific instructions or guides. In such cases, students are free to write about any aspect of their mathematics learning. Three such journals written by a secondary 2 student are shown in the following three figures: Figures 1, 2 and 3. The student created her own template for the journals and penned her thoughts periodically about her algebra lessons. The teacher's comments are also evident in the figures.



MY MATHEMATICS JOURNAL

NAME: Balpreet Kaur **CLASS:** 205

TOPIC: Expansion and Factorisation of "Algebraic Expressions" **DATE:** 19/1/03

Last year, we learnt the basics of Algebra. With more practice, last year's Algebra has become very easy to do. This year we will learn the Expansion and Factorisation of Algebraic Expressions, Quadratic Factorisation and Simultaneous Linear Equations. The Algebra this year will become more complex and sophisticated and hopefully more enjoyable but I doubt so.

So far, I have learnt the following expansions

– Multiplying binomials

for ex. $(x+8)(x+5)$ FOIL stands for First, Outer, Inner, Last

then we expand it using the FOIL method

$$\cancel{=x^2+40+8x} = x^2+5x+8x+40$$

$$= x^2+13x+40$$

However, I don't like this method and am not comfortable with it. I have my own method. $x(x+5)+8(x+5)$

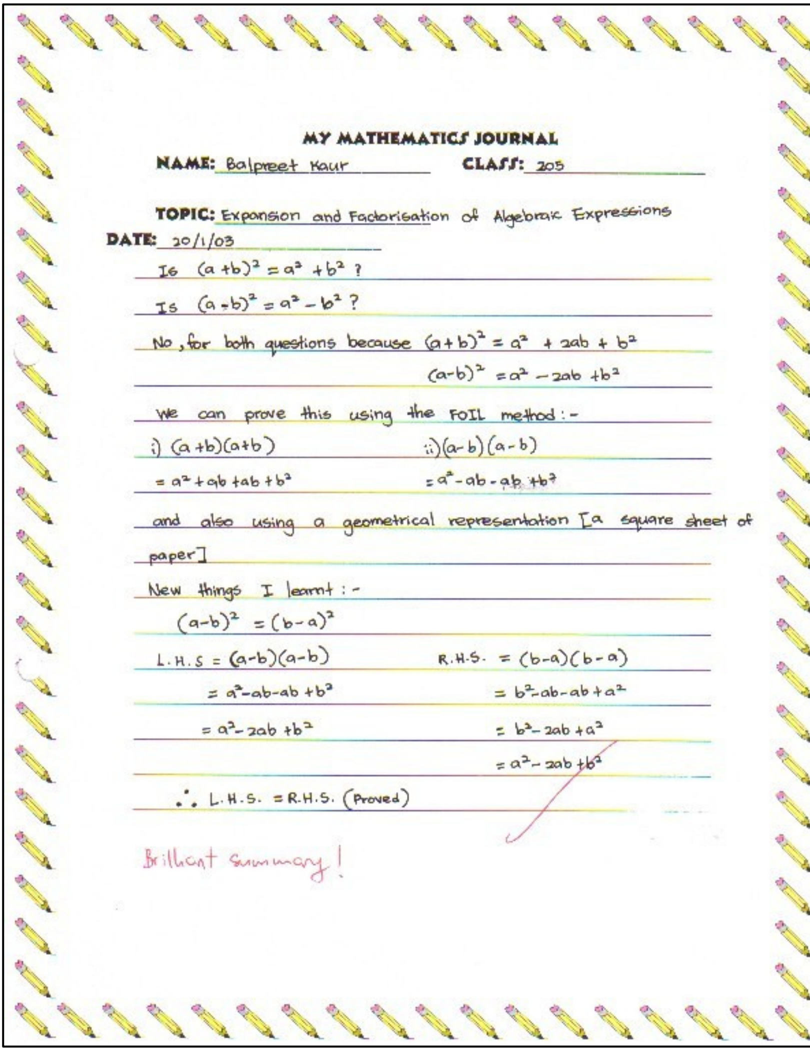
Step 1: $\underbrace{(x+8)(x+5)}_{x^2}$ Step 2: $\underbrace{(x+8)(x+5)}_{5x}$

Step 3: $\underbrace{(x+8)(x+5)}_{8x}$ Step 4: $\underbrace{(x+8)(x+5)}_{40}$

hahaha ...... sooo funny..... my method and the foil method are exactly the same. I just never realised it. But I still hate Acronyms. They further confuse me.

I find it very funny too! If you don't like FOIL, think of your own name. The acronym is not important at all :-)

Figure 1. Sample A of a free writing type of journal



MY MATHEMATICS JOURNAL

NAME: Balpreet Kaur **CLASS:** 205

TOPIC: Expansion and Factorisation of Algebraic Expressions

DATE: 20/1/03

Is $(a+b)^2 = a^2 + b^2$?

Is $(a-b)^2 = a^2 - b^2$?

No, for both questions because $(a+b)^2 = a^2 + 2ab + b^2$

$(a-b)^2 = a^2 - 2ab + b^2$

We can prove this using the FOIL method :-

i) $(a+b)(a+b)$ ii) $(a-b)(a-b)$

$= a^2 + ab + ab + b^2$ $= a^2 - ab - ab + b^2$

and also using a geometrical representation [a square sheet of paper]

New things I learnt :-

$(a-b)^2 = (b-a)^2$

L.H.S $= (a-b)(a-b)$ R.H.S. $= (b-a)(b-a)$

$= a^2 - ab - ab + b^2$ $= b^2 - ab - ab + a^2$

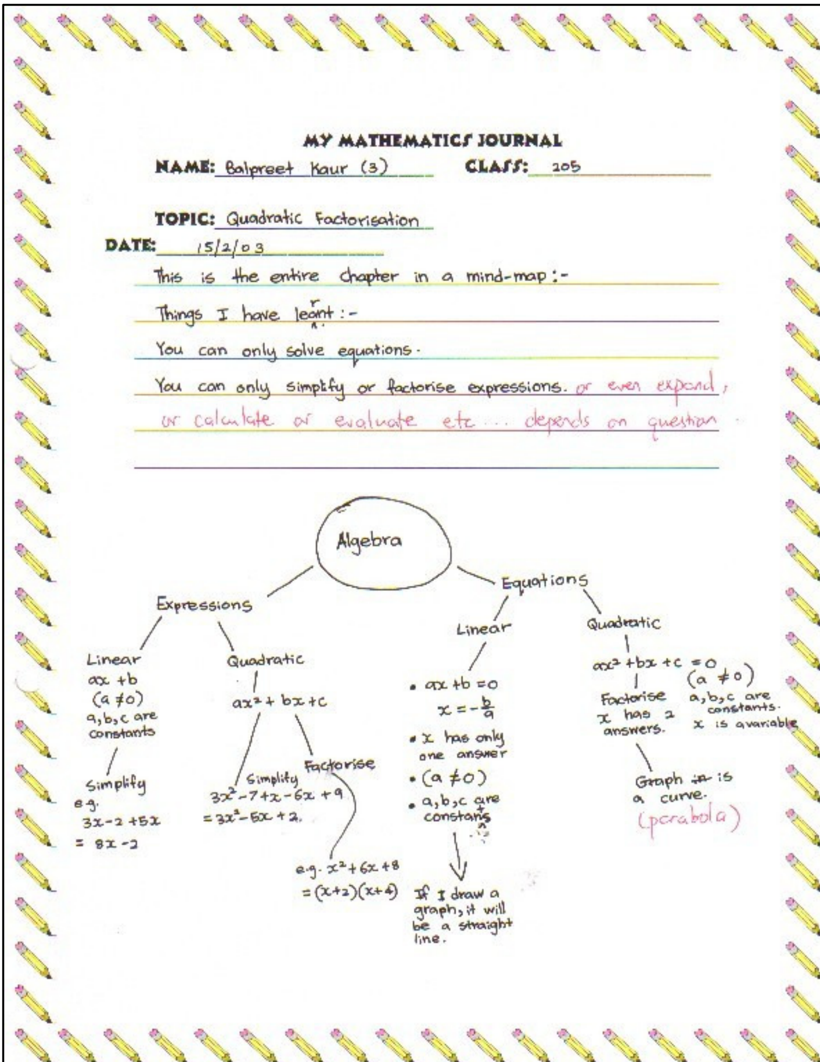
$= a^2 - 2ab + b^2$ $= b^2 - 2ab + a^2$

$= a^2 - 2ab + b^2$

$\therefore$ L.H.S. = R.H.S. (Proved)

Brilliant summary!

Figure 2. Sample B of a free writing type of journal


MY MATHEMATICS JOURNAL

NAME: Balpreet Kaur (3) CLASS: 205

TOPIC: Quadratic Factorisation

DATE: 15/2/03

This is the entire chapter in a mind-map :-

Things I have learnt :-

You can only solve equations.

You can only simplify or factorise expressions. or even expand, or calculate or evaluate etc... depends on question.

Figure 3. Sample C of a free writing type of journal

3.2 *Writing from a prompt*

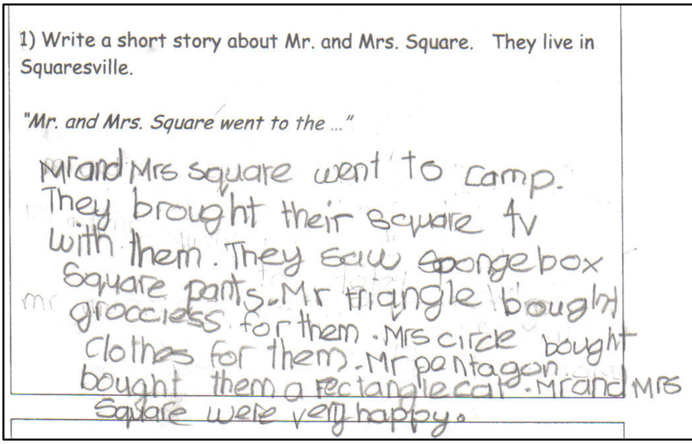
Alternatively, students may be given prompts to help them write journals. Through carefully chosen prompts, teachers can encourage students to connect new knowledge with prior knowledge and to develop new understandings about concepts. Kaur et al., (2004) and Kaur, Chan and Edge (2006) described three main categories of prompts, namely affective / attitudinal, mathematical content, and process and application. A prompt might be a sentence to complete, a question to respond to, or a quote to explain. The three categories of prompts follow.

Affect / Attitudinal (How do you feel?). Examples of prompts in this category are as follows:

- This year mathematics is …
- What I like most about mathematics …
- The most difficult part of mathematics is …
- What kind of mathematics figure are you? (circle, square, triangle, parallelogram, etc.) Explain;
- This is how I used mathematics this week, outside of school….

<u>Examples of journals in this category</u>
Primary 3 students were given the following prompt to write a short story about ***Mr and Mrs Square living in Squaresville***. Figure 4 shows a journal written by a student.

In the journal shown in Figure 4, the student wrote about things associated with squares since the characters were squares. Apparently the student drew on his or her personal knowledge such as a square TV, and the cartoon character Spongebox Squarepants that is a square-shaped creature. The student also named other characters using names of shapes such as Mr Triangle, Mrs Circle and Mr Pentagon to be consistent with the way Mr and Mrs Square are used. The composition, although short, paints a happy shopping occasion and a meeting of friends.


1) Write a short story about Mr. and Mrs. Square. They live in Squaresville.

"Mr. and Mrs. Square went to the ..."

Mr and Mrs square went to camp. They brought their square tv with them. They saw spongebox square pants. Mr triangle bought groccies for them. Mrs circle bought clothes for them. Mr pentagon bought them a rectangle car. Mr and Mrs square were very happy.

Figure 4. Sample journal depicting a story with a focus on shapes

A class of secondary 2 students were asked to complete the following: ***My worst experience with mathematics was*** …. Figure 5 show the response of one student in the class. This student appears to have found it difficult, when she was in primary 5, to accept that in mathematics, letters may be used instead of numbers to represent variables.

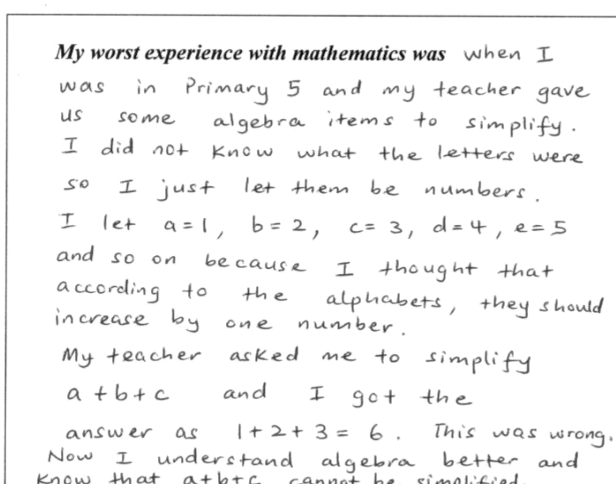

My worst experience with mathematics was when I was in Primary 5 and my teacher gave us some algebra items to simplify. I did not know what the letters were so I just let them be numbers. I let a=1, b=2, c=3, d=4, e=5 and so on because I thought that according to the alphabets, they should increase by one number. My teacher asked me to simplify a+b+c and I got the answer as 1+2+3=6. This was wrong. Now I understand algebra better and know that a+b+c cannot be simplified.

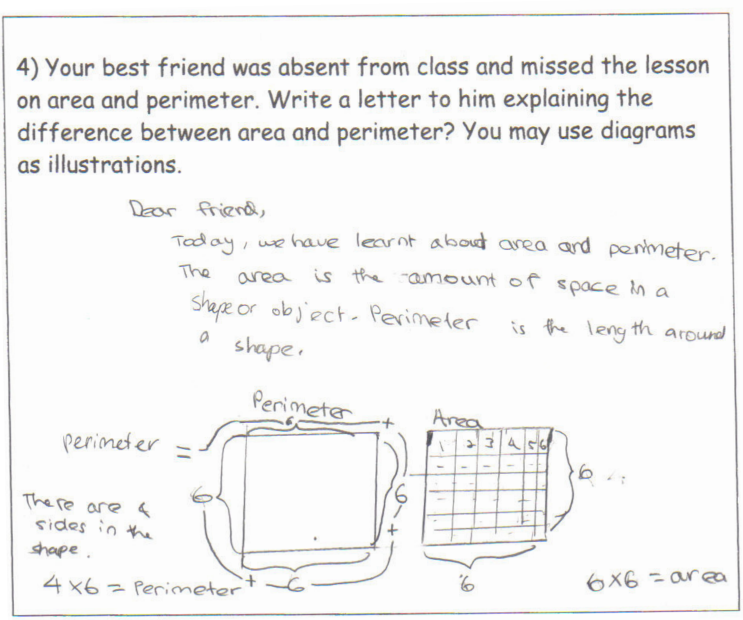
Figure 5. A student's journal on worst experience with mathematics

Mathematical content (What is it about?). Examples of prompts in this category are as follows:

- How would you describe a…?
- Write your own definition of …
- Write and solve a problem whose solution involves …
- Why can't you divide a number by zero?
- Compare and contrast….

Examples of journals in this category

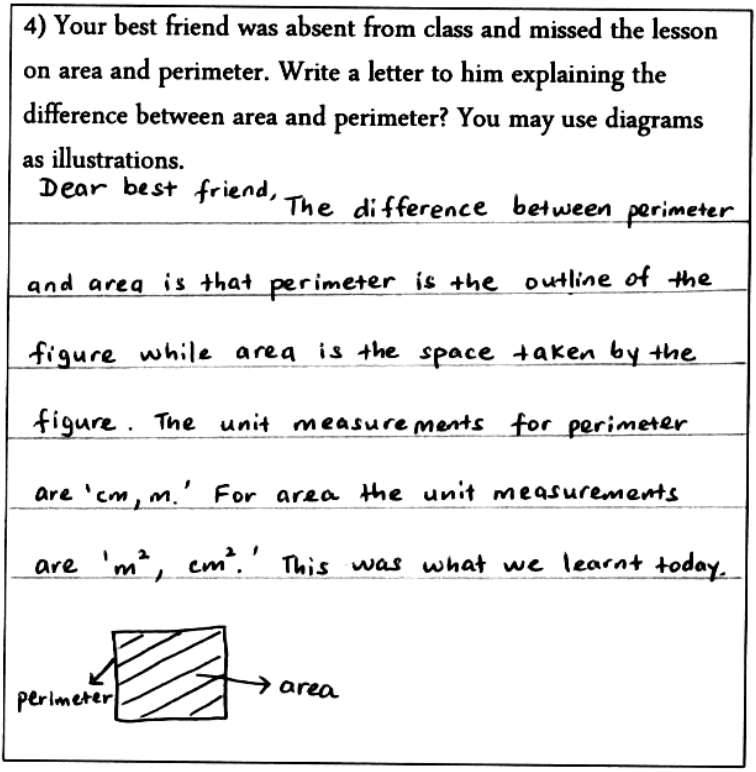
To determine if students had comprehended the difference between area and perimeter, they were asked to write a letter to an imaginary friend who had missed a lesson on area and perimeter. Figure 6 shows a journal entry of a Primary 4 student (student A).


4) Your best friend was absent from class and missed the lesson on area and perimeter. Write a letter to him explaining the difference between area and perimeter? You may use diagrams as illustrations.

Dear friend,
Today, we have learnt about area and perimeter. The area is the amount of space in a shape or object. Perimeter is the length around a shape.

Figure 6. Student A's journal on the difference between area and perimeter

From the journal, in Figure 6, it can be ascertained that student A has a good understanding of the two measurement ideas. The student is able to describe area as "the amount of space in a shape" and perimeter as "the length around a shape". The descriptions are enhanced with the illustrations. The square on the left is drawn and labeled with "6" on each side. The line drawn round the square shows that the perimeter is the sum of the lengths of the four sides. In addition, the statement "4 × 6 = perimeter" is written to show how the perimeter can be found. The square on the right is divided into smaller squares illustrating an understanding of the concept that leads to the formula: area = length × breadth. The statement "6 × 6 = area" is written to show how the area can be found. Figure 7 shows the journal entry of another Primary 4 student (student B).


4) Your best friend was absent from class and missed the lesson on area and perimeter. Write a letter to him explaining the difference between area and perimeter? You may use diagrams as illustrations.

Dear best friend, The difference between perimeter and area is that perimeter is the outline of the figure while area is the space taken by the figure. The unit measurements for perimeter are 'cm, m.' For area the unit measurements are 'm^2, cm^2.' This was what we learnt today.

Figure 7. Student B's journal showing the difference between area and perimeter

From Figure 7, it is apparent that student B was able to describe the difference between area and perimeter in a slightly different way from student A. Student B conveyed perimeter as the "outline of the figure" and area as the "space taken by the figure". The descriptions are aided by a diagram. The descriptions and diagram are complementary in showing what the student knew about the different ideas. The student also highlighted the unit of measurement for perimeter and area but did not show how perimeter or area of shapes may be computed.

Similarly to assess if students had understood the role of variables in algebraic expressions, a class of secondary 3 students were asked "Which is larger in magnitude? 5 + *n* or 5*n*, explain your answer clearly." Figure 8 shows a response of a student in the class. This student had a good understanding of the concept of a variable but had not considered all possible values for *n*.

MATHEMATICS JOURNAL

Name: [redacted] **Date:** [redacted]
Class: [redacted]
Topic: Algebra **Level: Secondary 1/2/3/4**

Which is larger in magnitude?
$5 + n$ or $5n$
Explain your answer clearly.

It depends on what n is.
If n is smaller than 1 or equal to 1, then 5 + n is larger than 5n. For example, if n is $\frac{1}{2}$, then 5 + n will be $5\frac{1}{2}$ and 5n will be $2\frac{1}{2}$.
If n is larger than 1, then 5n is larger than 5+n. For example, if n is 2, then 5+n will be 7 and 5n will be 10.
If n is 0, then 5+n will be larger than 5n as 5+n will be 5 and 5n will be 0.

Teacher's comments:
Good understanding of the concept of a variable n. However, you have not explored negative values of n and the value of n when 5 + n = 5n.

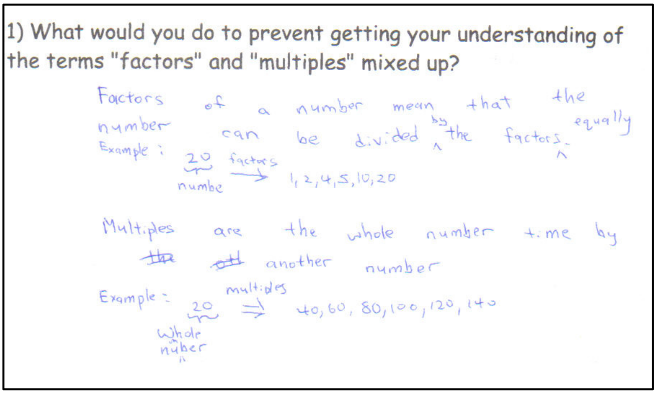
Figure 8. A student's journal on the comparison of two algebraic expressions

Process and Application. Examples of prompts in this category are:

- The most important part of solving a problem is …
- Do you use tables or diagrams when solving a problem? Why or why not?
- When I study for a test, I …
- Describe any computational procedure that you have invented;
- Write a letter to your teacher explaining to him / her how best you learn mathematics.

<u>Examples of journals in this category</u>

As pupils have the tendency to get some mathematical terms like *factors* and *multiples* mixed up, a class of Primary 4 students were asked to write about how they differentiated one from the other. Figure 9 shows one student's way of thinking in preventing the mixed-up. The student considered "factors of a number that can be divided by the factors equally". Although not eloquently phrased, the mathematical example shown helps to make sense of the description. The student used division as a test to see if it results in any remainder. The student was able to list all the factors of a particular number and knew that these numbers can be divided equally by the number. To know what multiples are, the student multiplied a number with another natural number. Again, the ability to use a mathematical example suggests what the student knew about multiples. In a sense, the student used "division" and "multiplication" in differentiating between the two terms.



1) What would you do to prevent getting your understanding of the terms "factors" and "multiples" mixed up?

Factors of a number mean that the number can be divided by the factors equally

Example : 20 numbe → factors 1, 2, 4, 5, 10, 20

Multiples are the whole number time by ~~the~~ ~~ot~~ another number

Example : 20 whole nuber ⇒ multides 40, 60, 80, 100, 120, 140

Figure 9. A student's journal depicting a way to differentiate factors and multiples

Figure 10 shows a journal entry written by a secondary 3 student in response to a writing prompt to elicit students' thinking processes when solving problems. The journal was submitted to the teacher electronically as part of work assignment during e-learning week.

MATHEMATICS JOURNAL	
Name: _XXX___	Date: _XXX_____
Class: _XXX___	Level: Secondary 3
Do you use tables or diagrams when solving a problem? Why or why not?	
I prefer to use tables or diagrams when solving a problem, where possible. This is because, visually representing a problem helps me to view the information provided by the problem and the unknowns I am required to find as a whole big picture, which can also be viewed as a composite of individual parts. Using the diagram, I usually derive a suitable systematic way of working with the information provided to arrive at an answer. Though, this may seem like a long-drawn process, it is actually a very fast one, as once you have visually represented the problem, you are no longer at a loss of how to get started. Spotting a solution from a visual representation is often very simple and clear-cut.	

Figure 10. A student's journal about use of tables and diagrams when solving problems

4 Rubrics for Grading Journals

Journal writing meant for the development of the affective domain of learners is often not graded. However, journals that are related to specific objectives of content, processes and applications may merit some form of grading as the feedback is relevant for both the teacher and the learner. As there is often no right or wrong answer to any journal entry, the use of scoring rubrics may be the most appropriate means of grading journals. Scoring rubrics provide at least two benefits in the evaluation process (Moskal, 2000). Firstly, they support the examination of the extent to which the specified criteria have been reached. Secondly, they provide feedback to students concerning how to improve their performances. Depending on the needs of the evaluator, an Analytic Scoring Rubric or a Holistic Scoring Rubric may be used.

4.1 *Analytic scoring rubric*

The Analytic Scoring Rubric allows for separate evaluation of areas such as:

- Mathematics content
- Organisation of ideas
- Expression

Figure 11 shows an Analytical Scoring Rubric from the Mathematics Assessment Guide (MOE, 2004b, p. 55).

Area	Score	How it was done
Mathematics Content	3	Showed in every instance, strong links between mathematics learning and daily life application. Used appropriate terms.
	2	Need some improvement
	1	Need to be significantly improved
Organisation of ideas	3	Very logical and systematic presentation
	2	Need some improvement
	1	Need to be significantly improved
Expression	3	Clear and coherent. Used appropriate diagrams.
	2	Need some improvement
	1	Need to be significantly improved

Figure 11. A journal writing task and an Analytical Scoring Rubric

4.2 *Holistic scoring rubric*

At times, it is not possible to separate an evaluation into independent areas. When there is an overlap between the criteria set for evaluation of the different areas, a Holistic Scoring Rubric may be preferable to an Analytic Scoring Rubric. Figure 12 shows a Holistic Scoring Rubric from Kaur, Chan and Edge (2006, p. 16).

Grade	Performance
A	• Response is complete and goes beyond expectations • Displays clear understanding • Ideas are clearly communicated • Complete and correct calculations are provided
B	• Response is complete • Displays understanding • Ideas are fairly well communicated • Work contains minor flaws in reasoning or minor errors in calculation
C	• Response is incomplete • Displays a lack of understanding • Ideas are not well communicated • Work contains numerous errors and reasoning is flawed
D	• No response or ideas are completely inadequate.

Figure 12. A Holistic Scoring Rubric

The rubric, shown in Figure 12, is applied to two samples of students' journals that address the prompt "the best way to find the sum of the interior angles of a pentagon is..." As part of the prompt, a pentagon is drawn for the students. Figure 13 shows student A's response to the prompt.

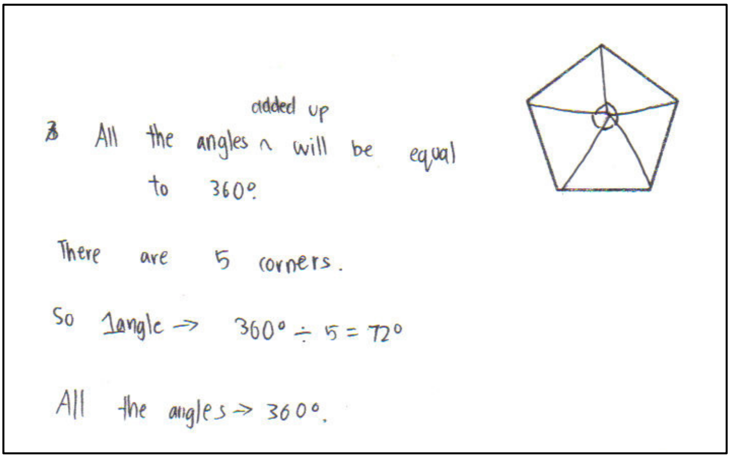


Figure 13. Student A's journal on how to find the interior angles of a pentagon

The student had interpreted the interior angles as the angles at the centre of the pentagon and drew lines from the vertices to the centre of the pentagon. The student referred to the five angles as “corners”. The student knew that since the sum of the 5 angles makes up 360^o, so each angle is found by dividing 360^o by 5 to get 72^o.

To apply the rubric, it is noted that the student did not quite know what are interior angles. Furthermore, 72^o as the value of an interior angle did not answer the prompt to find the sum of the interior angles. Hence, the student’s response was incomplete, but it showed partial understanding and warrants a C grade. Detailed qualitative feedback may be given by the teacher to the student suggesting that he failed to identify all the interior angles and how the angle with measure 72^0 may be used to work further towards the complete solution. Figure 14 shows student B’s response to the same prompt.

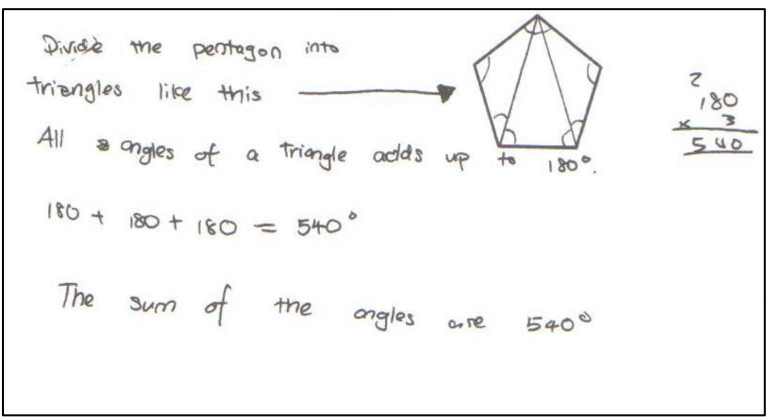


Figure 14. Student B’s journal on how to find the interior angles of a pentagon

Student B showed understanding of interior angles very clearly by marking out all the interior angles. To find the sum of the interior angles, the student divided the pentagon into three triangles. The student stated a property of triangles, i.e., the sum of the angles in a triangle is 180^o. Using the property the student found the sum of the interior angles of the pentagon.

According to the rubric, this response was complete and very clearly communicated. The student had used her geometrical knowledge correctly. The calculations were accurate and complete. Hence, the response of student B merits grade A.

5 Implementing Journal Writing in your Classroom – Potential Pitfalls

Like all forms of assessment, journal writing too may cause more harm than good if implemented mindlessly and carelessly. The followings are some pitfalls for teachers to take note of and avoid.

5.1 *The potential for teacher to hurt student's feelings*

Teachers must avoid criticizing what the students' write. They have to be positive, personal, accepting, encouraging and sensitive when responding to what students write in their journals. However this does not mean that teachers should avoid probing the students. It is improper for teachers to discuss journals of students in the teachers' room or quote from them during lesson time, especially if the journals are telling of some signs that are not favourable. However, the teacher may with the consent of the students concerned, occasionally share with the class very interesting or novel journals.

5.2 *Possible loss of instructional time to teach the syllabuses*

As the saying goes "water not drunk in moderation is harmful to the body", teachers have to be mindful of not going overboard with journal writing. A good planned programme is necessary to induct students into writing about their learning of mathematics and mathematical ideas. If the programme is well structured it will not bite significantly into instructional time. Also, the writing must be progressive. It should start with simple, short and focused tasks and progress to longer and more challenging (demanding the skills of analyses, synthesis and evaluation) tasks. Students may be asked to do their journal writing during lesson time or after school hours.

5.3 *Tremendous increase in the marking load of the teacher*

Certainly, when a teacher requires her students to write a journal after every lesson and grades all of them, her marking load would increase many folds. To avoid such a situation the teacher must:

- not go overboard with journal writing,
- be mindful that it is not necessary to grade all the journals all the time, and
- when grading journals, it is appropriate to focus on one or two areas rather than all possible areas.

5.4 *What to grade? Language or mathematics content*

When grading journals, teachers may be in a dilemma, as what to grade – the mathematics or the language of communication. Some teachers may be uncomfortable with the idea of grading writing and this is justified. Teachers should place emphasis on the mathematics content, organization of ideas and expression (MOE, 2004a and 2004b) when grading journals. Students can express themselves through diagrams, symbols and mathematical terminology. Teachers must not penalize students for errors in language of communication, i.e., grammar and punctuation.

6 Concluding Remarks

In this chapter, we have demonstrated that journal writing as an alternative mode of assessment for learning has many virtues. Bearing in mind the potential pitfalls, teachers are encouraged to engage their students in reflecting and writing about their mathematics learning, i.e., using journal writing to empower learning.

Acknowledgement

The authors thank Balpreet Kaur for giving them permission to showcase her journals in Figures 1, 2, and 3. The authors also thank all anonymous students whose work has provided the content for this chapter.

References

Bell, S. B., & Bell, R. N. (1985). Writing and mathematical problem-solving: arguments in favour of synthesis. *School Science and Mathematics*, *85*(3), 210-221.

Borasi, R. and Rose, B. J. (1989). Journal writing and mathematical instruction. *Educational Studies in Mathematics*, *20*, 347-365.

Burns, M. (2004). Writing in math. *Educational Leadership*, *62*(2), 30-33.

Chai, H. L. (2004). *The effects of using journal writing as an alternative assessment in a primary four mathematics classroom*. Unpublished MEd thesis. Singapore: National Institute of Education.

Clarke, D. J., Waywood, A., & Stephens, M. (1993). Probing the structure of mathematical writing. *Educational Studies in Mathematics*, *25*(3), 235-250.

Ernig, J. (1977). Writing as a mode of learning. *College Composition and Communication, 28*, 122-128.

Evans, C. S. (1984). Writing to learn in math. *Language Arts*, *61*, 828-835.

Kaur, B., Ang, C. T., Kang, Y. F .F., Ng, L. Y., Ng, T. C., Phua, C. G .R., Ibrahim, R., Seow, A. Y . J., & Tan, L. K. (2004). *Journal writing in the mathematics classroom (Secondary)*. Singapore: National Institute of Education.

Kaur, B., Chan, C. M. E., & Edge, D. (2006). *Journal writing in the mathematics classroom (Primary)*. Singapore: National Institute of Education.

Ministry of Education. (2000a). *Mathematics syllabus – Primary*. Singapore: Author.

Ministry of Education. (2000b). *Mathematics syllabus – Lower secondary*. Singapore: Author.

Ministry of Education (2004a). *Assessment guide to primary mathematics*. Singapore: Author.

Ministry of Education (2004b). *Assessment guide to lower secondary mathematics*. Singapore: Author.

Mok, M. C. M. (2010). *Self-directed learning oriented assessment: Assessment that informs learning and empowers the learner*. Hong Kong: Pace Publishing Ltd.

Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation*, *7*(3). Retrieved October 18, 2010 from http://PAREonline.net/getvn.asp?v=7&n=3.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

Ng, L. S. (2004). *Using a mathematics dictionary to improve students' performance in geometry*. Unpublished MEd thesis. Singapore: National Institute of Education.

Pugalee, D. K. (2005). *Writing to develop mathematical understanding*. Norwood, MA: Christopher-Gordon.

Shepard, R. G. (1993). Writing for conceptual development in mathematics. *Journal of Mathematical Behavior, 12*, 287-293.

Yeo, S. M. (2008). *Using journal writing and oral presentation as alternative assessment strategies in mathematics instruction: An empirical study in Singapore secondary schools*. Unpublished doctoral thesis. Singapore: National Institute of Education.

Chapter 6

Implementing Alternative Assessment in the Lower Primary Mathematics Classroom

YEO Kai Kow Joseph

Assessment is an integral part of the teaching and learning process. This chapter discusses issues related to assessment practices in the mathematics classroom. It also focuses on the importance of alternative assessment in the lower primary mathematics classroom. Amongst others, lower primary mathematics school teachers should consider implementing the following alternative assessment practices in their instructional process: practical tests, oral presentation, journal writing and open-ended tasks. Each type of alternative assessment practice is highlighted with examples followed by a short discussion.

1 Introduction

Educational and assessment policies come and go but the main aim of assessment will continue to inform teaching and learning. Assessment is expected to be among the most contentious issues in the mathematics classroom. Everyone has a view about assessment. It may mean different things to different educators and researchers. When educators and researchers respond to changes in assessment policy, they could be responding to a different facet of assessment. Over the last five years in Singapore, the mathematics curriculum was revised to place emphasis on reasoning, communications and connections; applications and modelling in addition to heuristics and thinking skills as processes that encompass the mathematics curriculum. As a result of this revision, there is a need to examine the instructional approach, the types of problems used and assessment approach in the primary mathematics classroom. The Ministry of Education (MOE) in Singapore announced in July 2010 that by year 2013, there will be no formal examinations for all primary one pupils (Davie, 2010). Instead of taking formal examinations, they will go through “bite-size formal assessment”. These bite-sized assessments inform the pupil, teacher and parents about the pupil’s areas of strength and areas to work on in his or her overall development. In other words, assessment will be continual, rather than limited to the mid-year and year-end examinations. With all these changes in the lower primary mathematics classroom, teachers must use assessment tools such as rubrics to assess and provide pupils with richer feedback on their development in both academic and non-academic areas (MOE, 2009). This is aligned with the established statement on the concept of assessment in mathematics that was given in the National Council of Teachers of Mathematics (NCTM) *Assessment Standards for School Mathematics*, which defined assessment as “the process of gathering evidence about a student’s knowledge of, ability to use, and disposition toward mathematics and of making inferences from that evidence for a variety of purposes” (NCTM, 1995, p.3). This strong sense of assessment informing instructional practice is also evident in Singapore’s Primary Education Review and Implementation (PERI) report (MOE, 2009). One of the PERI recommendations is to balance the acquisition of knowledge with the development of skills and values through increased use of engaging pedagogies, more holistic assessment to support pupils’ learning and development, and a stronger emphasis on non-academic aspects within the curriculum (Fu, 2010). Singapore schools are encouraged to explore alternative and more holistic forms of assessment at the lower primary levels to support greater balance between knowledge, skills and values acquisition.

Since the mid 1980s, educational researchers and teachers have suggested and practiced a wide range of alternative ways of assessing pupils’ learning to prevail over the shortfalls of the traditional paper-and-pencil test (Adam, 1998; Berenson and Carter, 1995; Clarke, 1997; Fan, Quek, Koay, Ng, Pereira-Mendoza, Yeo et al., 2008; MOE, 2004a and

2004b; Raymond, 1994; Richardson, 1988; Stacey, 1987; Stempien and Borasi, 1985). In the last ten years in Singapore, alternative assessment has also gained increasing attention from educational policy makers, administrators, researchers, and teachers, particularly since the early 2000s (Chow and Fan, 2005; Fan and Yeo, 2007; Fan, Quek, Koay, Ng, Pereira-Mendoza, Yeo et al., 2008; MOE, 2004a and 2004b; Seto, 2002; Yazilah and Fan, 2002). The PERI report also indicates that there is a need to "shift assessment practices away from an over-emphasis on assessment of learning as an end-outcome, especially at the lower primary levels and shape mindsets to view assessment as an integral part of ongoing processes to support learning" (MOE, 2009, p. 30).

In my work with lower primary mathematics school teachers in Singapore, I frequently hear concerns about implementing new assessment practices in mathematics classrooms. Anecdotal evidence suggests mathematics teachers understand the value of alternative assessment in assessing some instructional objectives more validly than the traditional mathematics test but this information may or may not be translated well into classroom practices. Lower primary mathematics school teachers may not feel confident about designing, implementing or judging pupils' work on alternative assessment tasks. There is a need to equip lower primary mathematics teachers with a set of newer alternative assessment practices to be integrated into their classroom instruction. This need is congruent with the PERI recommendations about assessments in that teachers must be aware of a range of pupils' ability and learning styles; be fair to all pupils and free from bias; as well as delineate and communicate assessment standards to pupils and parents.

The main purpose of this chapter focuses on implementing these principles in pragmatic ways. An important priority of the discussion is that the recommended alternative assessment practices should involve minimal disruption to the teaching process and not impose additional workload on the lower primary mathematics teachers. Rather, lower primary mathematics teachers are urged to be more judicious and watchful in their assessment practices. The ultimate purpose of classroom assessment is thus to facilitate and promote learning in the classroom. This chapter is divided into three sections. In the first section, I provide a relatively broad view of assessment practices in mathematics classrooms. In the second section, I provide four different alternative assessment practices that could be implemented in lower primary mathematics classroom. The chapter ends with a few concluding remarks.

2 Assessment Practices in Mathematics Classrooms

As assessment is a critical part of the teaching and learning process in classrooms, mathematics teachers need to keep abreast of new developments in assessment and be equipped with the necessary knowledge and skills in implementing various assessment practices. Traditionally, mathematics teachers have relied on paper-and-pencil tests to assess pupils' mathematics learning. Many of the mathematical items that teachers ask pupils to solve enable them to reproduce memorised procedures without thinking about why the processes work and what the answer means. The author recalls a test that a teacher gave to his grade two pupils to add the first twenty numbers 1 + 2 + 3 + 4 +…+ 20. Most mathematics teachers would administer similar test: first, pupils are asked to find the sum by adding up all the numbers in both horizontal and vertical formats, then by making 10 pairs of 21, and finally by any method that pupils wish. After this teacher had attended a workshop on open-ended problems, he added to his test the problem "List two sets of twenty numbers that have a sum of 210". Most pupils who were able to solve nearly all the standard addition problems correctly were unsuccessful on the new open-ended problem. This suggests that these pupils had only achieved instrumental understanding of solving addition problem. They had memorised procedures that allowed them to produce correct answers, but they were not able to show relational understanding of what the sum of a set of numbers really meant. It is apparent the first test focuses on static, discrete pieces of knowledge, often memorised but neither contextualized nor applied.

The traditional mathematics test possibly does not provide a comprehensive measure of pupils' ability. Traditional assessment techniques also make it difficult to develop inferences about pupils' learning that may be helpful in designing new approaches to improve pupils' learning. Webb (1992) claims that tests are important quantitative assessment tools, but they do not constitute the totality of assessment. Thus in this age of accountability, teachers need more varied information about their pupils' mathematical understanding and competence. One of the most important practices is for teachers to use effective classroom assessments that measure what pupils are learning and how they are learning it. Subsequently, teachers are expected to use the assessment information to plan and adjust instructional processes. In other words, classroom assessments should be integrated with the instructional process for mathematics teachers to understand and strengthen pupils' learning. These are the key principles of assessment **for** learning or formative assessment as encouraged by many assessment experts (Black and Wiliam, 1998; Shepard, 2000). Assessment **for** learning is a process by which assessment information is used and interpreted by teachers to adjust their teaching strategies for improvement in subsequent learning and teaching (Black, Harrison, Lee, Marshall and Wiliam, 2003; Black and Wiliam, 1998). Assessment **of** learning, commonly associated with national and school semester examination, is assessment which gathers evidence for summative judgement of pupils' performance.

It seems that support for, and understanding of, new assessment practices by teachers is inconsistent because of the lack of a well-defined set of assessment artifacts or practices that are readily transferrable across classroom, school and country contexts (Bennett, 2010; Maxwell, 2004). For example, even though there is strong belief and commitment by many teachers to implement formative assessment practices, they have found the implementation of such practices to be challenging (Carless, 2005; Dixon and Haigh, 2009; James, Black, McCormick, Pedder, and Wiliam, 2006). A major concern raised by teachers about the use of formative assessment methods relates to their perceived subjectivity. Just as assessment impacts pupils' learning and motivation, it also affects the nature of instruction in the classroom. There has been extensive recent literature which supports assessment as something to be integrated into the teaching process and not an activity that merely audits learning (Shepard, 2000). When assessment is integrated with instruction, it informs teachers about what activities and assignments will be most useful, what level of teaching is most appropriate, and how formative assessments provide diagnostic information. For instance, during instructional activities, teachers use formative assessment to know when to move on, when to ask more questions, when to give more examples, and what responses to pupils' questions are most appropriate.

Alternative assessment practices provide a more comprehensive picture of the pupil and provide more authentic information than traditional assessment practices which provide limited information about a pupil's understanding. In addition, alternative assessment may be appropriate for the assessment of non-academic domains because some of these domains cannot be assessed using pencil-and-paper tests. There are two related reasons for implementing alternative assessment in classroom practice. The first is that alternative assessment reveals at a very concrete level what the curricular objectives are. A second reason is alternative assessment offers teachers examples of how they can activate their pupils' thinking and learning. Other terms used for alternative assessment are: performance assessment, practical assessment, or authentic assessment (Burton, 1996; Clarke, 1996; Niss, 1993; Wiggins, 1989). Possible alternative assessment practices include practical tests, oral presentations, journal writing and open-ended tasks. One common aspect of these methods of assessment is that they represent "alternatives" to the more traditional paper-and-pencil test formats found in so many classrooms.

An important focus of alternative assessment is feedback and not the alternative assessment method used per se. It is possible to conduct alternative modes of assessment in classroom but if we do not provide feedback for learning, we only deceive ourselves that we are doing assessment **for** learning when we are indeed doing assessment **of** learning. The fundamental issue here is not the method – it is about providing feedback for follow-up actions. In view of this, there needs to be alignment of curriculum, teaching and assessment in order to address essential learning that needs to occur during meaningful classroom instruction and to address pupils' outcomes in a more comprehensive way. Second, there is a need to audit assessment practices for balance. The key word here is "balance"; neither excess nor neglect. The practice is not about having semester examinations or not having semester examinations – rather it is a search for a balance of the two types of assessment. There is also a need to build mathematics teachers' capacity to implement alternative assessment practices because such innovation is relatively new to many mathematics teachers who may not have received adequate formal assessment training in their pre-service teacher education programmes. Thus to make alternative assessment successful at the lower primary level, it is crucial that training is provided for lower primary mathematics teachers to realise that alternative assessment practices serve different but potentially valuable purposes.

3 Suggested Alternative Assessment Practices for the Lower Primary Mathematics Classroom

Although paper-and-pencil mathematics tests may be used for specific purposes, there is a need to explore alternative assessment practices to assess other learning outcomes. The alternative assessment practices proposed in this chapter include those suggested in the Ministry of Education of Singapore assessment guidelines for instruction at the primary school levels (MOE, 2004a) as well as recommendations from research in mathematics education.

Among others, lower primary mathematics school teachers should consider implementing the following alternative assessment practices in their instructional process: practical tests, oral presentation, journal writing, and open-ended tasks. All these alternative assessment practices, which can be easily implemented in the lower primary mathematics classrooms, are described below.

3.1 *Practical tests*

The advantages of practical tests include the provision of short-term learning goals, enhanced motivation as well as instant and explicit feedback (Clarke, 1992). In practical tasks pupils are expected to use manipulatives, materials and instruments to deduce mathematical principles. A practical test is also an effective assessment tool and has a high degree of assessment validity as the skills are assessed in practice in the manner in which they have been learnt through the use of hands-on experiences. The use of everyday manipulatives, such as money, matchsticks, crayons, cubes and beads, give pupils a realistic view of the things around them since these assist pupils to better understand the context of the question. This would encourage pupils to be more aware of their surroundings and to observe everything as an opportunity for learning. It reminds pupils that mathematics is related to our daily life as they have to handle real objects such as weighing scales and measuring cups. Pupils are required to show the processes involved in working out the test items using the appropriate manipulative. These processes enable a fair assessment of pupils' understanding and therefore allow pupils to gain good learning experience through these hands-on activities.

Many topics in the lower primary mathematics curriculum lend themselves to practical tests, for example, Ordinal Numbers, Patterns, Length, Mass, Volume and Shapes. For primary one pupils who are new to the school system, practical tests would ease them into a more structured learning environment. This can be conducted by using manipulatives to engage pupils, assessment checklists to monitor their progress, and bite-sized practical tests to assess their understanding. As practical tests will be the initial experience for lower primary school pupils, it will be appropriate to focus on closed items. They can be assessed through closed items like: measuring and comparing the lengths and masses of two or more objects in non-standard units. Practical test items, such as using a kitchen scale to determine the mass of an item or using a ruler to measure length, require pupils to demonstrate that they have mastered estimation and measuring skills using mathematical tools. Other skills that the pupils might develop include observing and describing, looking for regularities, explaining and predicting measurements. More examples of practical test items include:

1. Putting objects on the balance to find out which object is heavier than the tube of glue.
2. Using paper clips to measure the length of the mathematics textbook on the table.
3. Using the measuring tape to measure the length, breadth and height of the box on the table.

4. Providing three bags of marbles, A, B and C and kitchen scale to find out the mass of the bags. Pupils arrange the bags in order of mass, starting from the lightest to the heaviest.
5. Providing two containers, A and B, filled with water to compare volume by pouring the water into the two beakers to find out which container holds more water.

Such practical tests can be administered in a circuit format, in which pupils take turns in visiting stations, carry out the investigation, and record their answers on a worksheet. Mathematics teachers need to be aware that a lot of time is needed to prepare the materials and to set up the stations. Alternatively, pupils can be assessed individually using the same format as the English oral examination where pupils are assessed individually outside the classroom. This one-to-one format enables the mathematics teacher to cater to each pupil's need. Pupils are not rushed and are given ample time to use the manipulatives to obtain their answers. However, we need to be vigilant when implementing practical tests for young children. For example, testing conditions need to be well-controlled because unreliable measurements will indicate incorrect differences between pupils and between classes. We need to be mindful that such small changes in materials, manipulatives and measuring instruments might destroy the reliability of a measurement. Nevertheless, by gaining experience using practical tests, teachers can gather more reliable data.

3.2 *Oral presentations*

Oral presentations enable pupils to give solutions verbally and the process of interaction between a teacher and pupils facilitates sharing of thoughts and clarification of understanding. One of the important aims of oral presentations in mathematics classroom is to create an opportunity for the teacher to listen to what the pupils are saying about their thinking about mathematics, how they communicate mathematically and their understanding of mathematics using their own words. In addition, according to the *Communication Standard for Grades 6-8* by the National Council of Teachers of Mathematics (NCTM), "teachers using oral presentation tasks must provide opportunity for pupils to think through questions and problems; express their ideas; demonstrate and explain what they have learnt; justify their opinions; and reflect on their understanding and on the ideas of others" (NCTM, 2000, p 272). There are two main benefits of using oral presentations. First, the teachers can gather information about their pupils' learning of mathematics and use this information to direct instructional process. The pupils' can develop communication skills. Teachers need to be aware that opportunity for pupils to be involved in active and meaningful verbal communication is a critical process for their learning and knowledge acquisition (Fan and Yeo, 2007). They also need to give pupils necessary guidance (including clarity of expectations), especially at the early stage, and in particular, create an encouraging classroom environment for pupils to engage themselves in such communication. As lower primary school pupils are still young, it is necessary to structure the oral presentation tasks so that the pupils and teachers are engaged. Examples of oral presentation tasks include:

1. Pupils' previous writing tasks on their learning reflection or perceptions.
2. Pupils' solutions to non-routine problems.
3. Pupils' previous writing tasks about learning of mathematical concepts.
4. A chosen idea that is pre-agreed before discussion.
5. To share their problem-solving behaviours.
6. To share results or findings of a learning journey through mathematics trial.
7. To present pair or group work activity.

Pupils should be given the chance to communicate their thinking to other pupils in about two to three minutes. With young children, it is necessary to encourage them to explain and clarify their thinking verbally.

3.3 *Journal writing*

Journal writing offers pupils opportunities to reflect on their learning by writing about their thoughts and feelings about the mathematics they are learning. Pupils keep reflective accounts of their mathematics learning and processes of understanding from which the teacher may grade the quality of their task. Similar to oral presentation, journal writing can be a valuable technique to further develop and enhance pupils' mathematical thinking and communication skills in mathematics. In other words, journal writing can also assist pupils to learn how to communicate mathematically when they try to explain what they have learnt. This may also help them to clarify their own understanding (Stempien and Borasi, 1985). Journal entries in mathematics provide opportunities for pupils to self-assess what they have learned. When pupils make an entry into a mathematics journal, it becomes a record of the experience received from the specific mathematics lesson or problem-solving activity. The pupil has to think about what he or she has done in order to communicate it in writing. When reading through the journal entries, the teacher decides if further review is required. It is best not to begin by having pupils write about unfamiliar mathematical ideas. First get them used to writing in a mathematics class. The teacher can begin with affective and open-ended questions about pupils' feelings. The following are some examples of mathematics journal prompts that might help pupils start their journal writing.

1. The things I want to know about in mathematics are …
2. I knew I was right when......
3. I wish I knew more about......
4. Were you frustrated with this problem? Why or why not?
5. What decisions had to be made when solving this problem?
6. Is mathematics your favourite subject? Why or why not?

Once pupils have become used to writing about their attitudes and feelings toward mathematics in their journals, they are ready to write about simple, familiar mathematics concepts. It is critical not to make the writing too difficult by asking lower primary school pupils to write about unfamiliar mathematics ideas. Using writing to review familiar mathematics ideas will increase confidence and skill in writing as well as revisit important mathematical concepts and processes. The following examples of mathematical journal prompts assist pupils to revisit important mathematical concepts and processes:

1. Explain in your own words what addition means.
2. Explain what is most important to understand about fractions.
3. What would happen if you missed a step in solving a problem? Why?
4. How many times did you try to solve the problem? How did you finally solve it?
5. The thing you have to remember with this kind of problem is........
6. What method did you use to solve this problem and why?

In addition, Waywood (1992) proposed three forms of journal writing: recount, summary, and dialogue. In the recount approach, pupils write what they have observed in their lessons. In the summary approach, pupils review what they have learnt during their lessons. In the dialogue approach, pupils elaborate on what they have learnt. Waywood also illustrated how to assess pupils' journal writing using a scoring rubric.

In the use of mathematics journals, pupils are required to express their understanding through drawing, mathematical formulae or words. It is thus useful to evaluate pupils' understanding and create opportunity for teachers to provide feedback to the pupils through journal writing. Besides, teachers can utilise journal writing as a formative assessment where they learn more about their pupils' learning difficulties or misconceptions from their journals and then proceed to remediate the situation (Miller, 1992). This form of journal writing should not be graded, otherwise the pupils may pretend that they understand everything in the mathematics lesson.

3.4 *Open-ended tasks*

Open-ended tasks elicit a range of responses from pupils including a chance for pupils to show all that they know about the relevant content. The purpose of open-ended tasks is to provide pupils with the opportunity to communicate their understanding in depth. Open-ended items "have more than one answer and/or can be solved in a variety of ways" (Moon and Schulman 1995, p. 25). In addition to producing a solution, pupils must also explain their solution process and justify their answer. According to De Lange (1995), a task that is open for pupils' process and solution is a way of motivating pupils' high quality thinking. Furthermore, Sullivan and Lilburn (2005) argue that open-ended tasks are exemplars of good questions in that they advance significantly beyond the surface. Specifically, open-ended tasks are those that require pupils to think more intensely and to provide a solution which involves more than remembering a fact or repeating a skill. Open-ended tasks offer opportunities for pupils to reveal their mathematical thinking, reasoning processes as well as problem-solving and communication skills. It is an attempt to make assessment more of an open process that will benefit both teachers and pupils. Although it is vital to assess pupils' mastery of mathematical skills, it is also essential to assess their conceptual understanding of mathematics. Often, just a little twist on the items we typically use in assessing our pupils can yield the assessment intent. Consider the following open-ended tasks for the lower primary levels:

1. Write down five whole numbers between 178 and 202.
2. List five 3-digit numbers that have digit 6 in the tens place.
3. List two sets of five numbers that have a sum of 100.
4. Draw a shape where the sum of all the sides is 36 cm.
5. Gilbert and Hazel have 40 postcards. After Gilbert gives a few postcards to Hazel, how many postcards does Hazel have? Explain your answer.
6. Draw a triangle. Write a number in the centre of the triangle. Write three numbers in the corners of the triangle that add up to the number in the centre. Now challenge yourself by choosing

greater numbers. Draw and write as many triangles and numbers as you can.

Such open-ended tasks emphasize the importance of thoroughly understanding concepts and carefully communicating mathematics, rather than short-answer items that ask pupils to simply practice rote memorization of mathematical facts. In lower primary mathematics classroom, after the teacher has taught some mathematical concepts or skills, the open-ended task is intended to elicit pupils' understanding of the concepts or ability to use the skills. These open-ended tasks expect pupils to generate examples that fit certain criteria and enable teachers to get a better vision of pupils' understanding of mathematical topics. Pupils need to develop their own methods for getting the right answers. One criterion for a good open-ended task is that it will elicit responses that are amenable to partial credit according to some established rubric. One interesting development in assessment in recent years has been the use of scoring rubrics with a greater emphasis on making more holistic judgments on pupil's work, with less emphasis on counting up "rights and wrongs". Rubrics are brief outlines that describe the content and quality of work needed to achieve a specific grade in an open-ended task and enable the teacher assessing the piece of work to determine the evidence of pupils' understanding and mathematical communication. Such an approach would focus on "big ideas", rather than simply facts and procedures.

These four alternative assessment practices exemplify how teachers and pupils could benefit from implementing alternative assessment in the lower primary mathematics classroom. The different alternative assessment practices highlight the different learning experiences that pupils will gain when they work on diverse types of tasks. This is only possible when the assessment tasks that teachers use in their classrooms go beyond computation and rote algorithms. The four alternative assessment practices are just the first steps towards making the use of alternative assessment in the classroom a meaningful one where emphasis is on the process (reasoning and thinking) rather than the product (final answer).

4 Concluding Remarks

This chapter has examined four ways in which mathematics assessments at the lower primary level might encourage more flexible, active and mindful learning. These ways may shift the focus of our present assessment that is highly dependent on paper and pencil to authentic assessments. This means that assessment tools can be build around the mathematical task that would enable teachers to gather evidence of pupil learning and use such evidence to further improve lessons. Implementing alternative assessment practices is only the first step; they are only valuable when teachers can use the assessment information to improve pupils' learning. This implies that teachers use the assessment information to change curriculum and instruction, so that what they teach and how they teach enhances what and how pupils learn. There should also be a balance between assessments that provide feedback and do not count towards a final grade and assessments that are graded to check for mastery of learning. Furthermore, teachers need to spread out alternative assessments appropriately so that their young pupils are not overwhelmed. The most effective teachers are likely to be those who approach assessment as an opportunity for pupils to show what they know and can do.

There are, however, numerous challenges if alternative assessment is to become a reality in the classroom. Undoubtedly it will require highly competent teachers who have mastered the complexity of the lesson where the pupils are continuously being actively engaged in constructing and applying the mathematical ideas and skills. Teachers' skills and knowledge are important to consider as well as the level of support and resources from the school administration. One main challenge therefore is to develop teachers' skills, knowledge and attitude to implement alternative assessment in the lower primary mathematics classroom. While formal training through workshops and seminars may be able to impart new knowledge to teachers, more discussions between teachers and mathematics educators would be required to bridge the gap between theory and practice. Teachers need support to acquire the skills and confidence to implement alternative assessments at the classroom level.

References

Adam, T. L. (1998). Alternative assessment in elementary school mathematics. *Childhood Education, 74*(4), 220-224.

Bennett, R. (2010). *A critical look at the meaning and basis of formative assessment*. Cambridge: Cambridge Assessment Seminar.

Berenson, S. B., & Carter, G. S. (1995). Changing assessment practices in science and mathematics. *School Science and Mathematics*, *95*(4), 182-186.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice.* Buckingham: Open University Press.

Black, P., & Wiliam, D. (1998). Inside the black box-raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139-148.

Burton, L. (1996). Assessment of mathematics: What is the agenda? In M. Birenbaum & F.J.R.C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes, and prior knowledge* (pp. 31-62). Dordrecht, NL: Kluwer.

Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, *12*(1), 39-54.

Chow, Y. P., & Fan, L. (2005, May). *Impact of open-ended problem solving as an alternative assessment on secondary one mathematics students*. Paper presented in the CRPP/NIE International Conference on Education: Resigning Pedagogy: Research, Policy, Practice. Singapore.

Clarke, D. J. (1992). Activating assessment alternatives in mathematics. *Arithmetic Teacher, 39*(6), 43-48.

Clarke, D. (1996). Assessment. In A. J. Bishop, K. Clement, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 327-370). Dordrecht, NL: Kluwer.

Clarke, D. (1997). *Constructive assessment in mathematics: Practical steps for classroom teachers*. Berkeley, CA: Key Curriculum Press.

Davie, S. (2010). Elephant in the classroom. *The Straits Times*, (p. A2), July 22, 2010.

De Lange, J. (1995). Assessment: No change without problems. In T. A. Romberg (Ed.) *Reform in school mathematics and authentic assessment* (pp. 87-173). New York: Suny Press.

Dixon, H., & Haigh, M. (2009). Changing mathematics teachers' conceptions of assessment and feedback. *Teacher Development*, *13*(2), 173-186.

Fan, L., Quek, K. S., Koay, P. L., Ng, C. H. J. D., Pereira-Mendoza, L., Yeo, S. M., Tan-Foo, K. F., Teo, S. W. & Zhu, Y. (2008). *Integrating new assessment strategies into mathematics classrooms: An exploratory study in Singapore primary and secondary* *schools* (Final Research Report). Singapore: Centre for Research in Pedagogy and Practice, National Institute of Education.

Fan L., & Yeo S. M. (2007). Integrating oral presentation in mathematics teaching and learning: An exploratory study with Singapore secondary students. *The Montana Mathematics Enthusiast*. USA. Monograph 3, 81-98.

Fu, G. (2010). *Opening address at the Primary Education Review and Implementation (PERI) holistic assessment seminar 2010*. Retrieved October 11, 2010, from http://www.moe.gov.sg/media/speeches/2010/07/13/peri-holistic-assessment-seminar 2010.php.

James, M., Black, P., McCormick, R., Pedder, D., & Wiliam, D. (2006). Learning how to learn, in classrooms, schools and networks: Aims, design and analysis. *Research Papers in Education*, *21*(2), 101-118.

Maxwell, G. S. (2004). *Progressive assessment for learning and certification: Some lessons from school based assessment in Queensland.* Paper presented in the Third Conference of the Association of Commonwealth Examination and Assessment Boards, Nadi.

Miller, D. L. (1992). Teacher benefits from using impromptu writing prompts in algebra classes. *Journal for Research in Mathematics Education, 23*, 329-340.

Ministry of Education. (2004a). *Assessment guides to primary mathematics*. Singapore: Author.

Ministry of Education. (2004b). *Assessment guides to lower secondary mathematics*. Singapore: Author.

Ministry of Education. (2009). *Report of the primary education review and implementation committee.* Singapore: Author.

Moon, J., & Schulman, L. (1995). *Finding the connections: Linking assessment, instruction, and curriculum in elementary mathematics.* Portsmouth, N.H: Heinemann.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics* . Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

Niss, M. (1993). *Investigations into assessment in mathematics education, an ICMI Study*. Dordrecht, NL: Kluwer.

Raymond, A. M. (1994). Assessment in mathematics education: What are some of the alternatives in alternative assessment? *Contemporary Education*, *66*(1), 13-17.

Richardson, K. (1988). Assessing understanding. *Arithmetic Teacher*, *35*(6), 39-41.

Seto, C. (2002). Oral presentation as an alternative assessment in mathematics. In D. Edge & B. H. Yeap (Eds.), *Mathematics education for a knowledge-based era* (Vol. 2, pp. 33-39). Singapore: Association of Mathematics Educators.

Stacey, K. (1987). What to assess when assessing problem solving? *Australian Mathematics Teacher*, *43*(3), 21-24.

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher, 29* (7), 4-14.

Stempien, M., & Borasi, R. (1985). Students' writing in mathematics: Some ideas and experiences. *For the Learning of Mathematics*, *5*(3), 14-17.

Sullivan, P. & Lilburn, P. (2005). *Open-ended maths activities: Using 'good' questions to enhance learning*. Melbourne: Oxford University Press.

Waywood, A. (1992). Journal writing and learning mathematics. *For the Learning of Mathematics, 12*(2), 34-43.

Webb, N. L. (1992). Assessment of students' knowledge of mathematics: Steps toward a theory. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 661-683). New York: Macmillan.

Wiggins, G. (1989). A true test: Towards more authentic and equitable assessment. *Phi Delta Kappan, 76*(9), 703-713.

Yazilah, A., & Fan, L. (2002). Exploring how to implement journal writing effectively in primary mathematics in Singapore. In D. Edge & B. H. Yeap (Eds.), *Mathematics education for a knowledge based era* (Vol. 2, pp. 56-62). Singapore: Association of Mathematics Educators.

Chapter 7

Open-Ended Tasks and Assessment: The Nettle or the Rose

David J. CLARKE

The focus of this chapter is the use of open-ended tasks in mathematics assessment. It is argued that the international mathematics education community has had access to a wide range of open-ended tasks for at least 20 years. The instructional utility of such tasks has been variously realised in different school systems around the world, however the role that open-ended tasks might play in assessment has not been explored sufficiently. Research findings are reported to highlight issues associated with the inclusion of open-ended tasks for the assessment of mathematics. Both positive and negative considerations are identified and it is argued that the assessment potential of open-ended tasks will only be achieved where teachers, schools and school systems are prepared to systematically address each consideration in developing programs of mathematics assessment that accommodate the complex array of behaviours that make up sophisticated mathematical practice and thinking. We have the tools but lack the assessment structures through which to employ them to model, monitor and inform valued mathematical performances. Recent curriculum innovations in Australia, China, Korea and Singapore will only achieve the intended transformation of classroom practice (and student outcome) if they can be matched equally by visionary assessment schemes. Open-ended tasks could be the key to such sophisticated assessment.

1 Introduction

> What we call the beginning is often the end
> And to make an end is to make a beginning,
> The end is where we start from. (Eliot, 1950, p.42)

The enthusiasm with which the use of open-ended and non-routine problems are being advocated for instructional and assessment purposes in mathematics classrooms needs to be tempered by consideration of existing research evidence. This is not intended to contest the value of open-ended tasks, but rather to ensure that their use is suitably informed by research and therefore most likely to prove effective. Recent mathematics curricula internationally offer a vision of sophisticated mathematical activity as the goal of school mathematics. It remains to be determined how we might reach this goal and, as importantly, how we will know when we have arrived.

The testing of skills and facts as an appropriate measure of mathematical learning depends on a conception of mathematical knowledge as hierarchical and discrete. It is now well-established that reliance on testing as the sole form of mathematics assessment is inappropriate: misrepresenting mathematics, at odds with contemporary curricula, misleading in the information it provides teachers, and potentially destructive in its effects on some learners (Clarke, 1992a and 1996). The article by Shepard (1991) titled, "Psychometricians' Beliefs about Learning" proposed that the disputes of the testing community could be explained in terms of differences in the beliefs about learning held by the various educational measurement specialists. In particular, Shepard argued that the beliefs of many psychometricians derive from an implicit behaviourist learning theory in flagrant contradiction with evidence from cognitive psychology. The need for sophisticated assessment tools capable of eliciting the complex performances now recognised as constituting mathematical practice and learning is captured neatly in Shepard's provocative question, "But what if learning is not linear and is not acquired by assembling bits of simpler learning" (Shepard, 1991, p. 7).

For more than 20 years, mathematics curricula have attempted to articulate this more complex vision of mathematical thinking and practice; from early examples such as *Realistic Mathematics Education* in the Netherlands (De Lange, 1987) and the American *Curriculum and Evaluation Standards* (NCTM, 1989) to more recent curricula. But such curricular aspirations demand an alternative vision of assessment. The realization of such a vision requires tasks that provide pupils with the opportunity to engage in mathematical activity in a variety of forms and contexts, and at different levels of sophistication.

Various initiatives in the UK, Australia and the USA have attempted to employ mandated assessment as a catalyst for systemic curricular change. Such initiatives sought to model mathematical thinking and practice through multi-component assessment systems. This modeling was particularly evident in the "Investigations" developed by the California Assessment Program (CAP, 1989 and 1991), and in the "Challenging Problems" and "Investigative Projects" of the Victorian Certificate of Education (see Barnes, Clarke, and Stephens, 2000). "Rich" assessment tasks such as these have a "problem solving" and "open-ended" character, in the sense in which these terms are generally used. The student responses elicited by such tasks are much more complex than those arising from conventional test items. The inferences that might be drawn legitimately from student performance on such open-ended tasks require research substantiation (Clarke, 1996; Clarke and Sullivan, 1992).

The OECD *Programme for International Student Assessment* (PISA) has attempted to give assessment recognition to the situated nature of mathematics activity to a greater extent than in the *Trends in International Mathematics and Science Study* (TIMSS). The attempt within international student achievement initiatives such as PISA to honor the situatedness of mathematical activity within an international testing instrument is wholly commendable. Of course, this same situatedness renders attempts at cross-curricular measurement of student mathematical performance somewhat problematic (see Clarke, 1996). The implicit recognition that mathematics can only be assessed "in use" and that such use implies a context, reflects the underlying assumptions of the Dutch *Realistic Mathematics Education* curriculum (De Lange, 1987), among others. The consequences of integrating such a perspective into an instrument intended to measure student mathematics achievement internationally can be seen in the observation that "national rankings on TIMSS and PISA differ substantially" (Törner, Schoenfeld, and Reiss, 2007, p. 353). More difficult to model in a test format such as those employed in TIMSS and PISA (and in many summative school mathematics examinations) are the forms of sophisticated mathematical activity, variously grouped together as "mathematical problem solving" or "mathematical thinking" or "working mathematically." International developments regarding these types of mathematical performance are usefully summarised in the special issue of *ZDM: The International Journal on Mathematics Education* titled "Problem Solving Around the World: Summing Up the State of the Art" (Törner, Schoenfeld, and Reiss, 2007). The commitment to honour such forms of sophisticated mathematical thinking in assessment activities may provide the strongest argument in favour of the use of open-ended tasks in assessment.

In line with recognition of the inadequacy of conventional testing as a measure of mathematics learning is the happy recognition that assessment need not be exclusively summative (Black and Wiliam, 1998). The valuing of formative classroom-based assessment in mathematics provides a fertile setting for the use of open-ended tasks in both instruction and assessment, since the premises for the advocacy of an increased role for formative assessment include the potential utilization of rich classroom activities for the simultaneous pursuit of instructional and assessment purposes.

While research suggests that such tasks may have instructional value (for instance, Clarke, Breed, and Fraser, 2004; Sullivan and Clarke, 1988 and 1991b), the assessment value of such tasks is less clear (Clarke and Sullivan, 1992; Clarke, Sullivan, and Spandel, 1992). Significant funds and personnel have been committed to the development and implementation of assessment systems employing such tasks. One purpose of this chapter is to ensure that the international mathematics education community subjects the use of open-ended tasks for assessment purposes to rigorous scrutiny in order that such tasks might usefully advance the contemporary curricular agenda. Significant curricular initiatives are being undertaken in Australia, China, Korea, and Singapore and these initiatives require the development and implementation of assessment tools commensurate with the sophistication of the new curricula.

2 Modelling Mathematical Thinking - Thesis

> When the tongues of flame are in-folded
> Into the crowned knot of fire (Eliot, 1950, p. 44)

The recognition that school mathematics has misrepresented mathematical activity provided some of the motivation for the call for "authentic assessment". At the heart of the matter lies the conflict of paradigms between a conception of mathematics as a catalogue of distinct and disjoint skills assessed by targeted test items and a conception of mathematics as a network of behaviours which are not acquired hierarchically, are not practised independently, and cannot meaningfully be assessed in isolation from each other. The challenge for mathematics instruction and assessment is to portray, to model, to stimulate, to facilitate, to recognize, and to acknowledge the complexity and the power of mathematical thinking in the context of school classrooms.

A wide variety of instructional activities have been developed to model this complexity, and to locate tasks in identifiable contexts. The Australian Mathematics Curriculum and Teaching Program (MCTP), the British materials produced by the Shell Centre, the adaptation in Wisconsin of materials from the Netherlands (Mathematics in Context), and the Californian Interactive Mathematics Project (IMP) all represent quite different attempts to engage students in purposeful mathematical activity in meaningful contexts. The goals of such materials are ambitious and include: more effective schema development; increased generalisability (transfer) of specific skills; and the development of higher order problem solving and reflective capabilities. From this standpoint, authentic assessment must offer students the opportunity to attempt tasks of varied complexity and contextual diversity. It should not be possible to identify a task in isolation as instructional or evaluative per se, but only through an examination of the teacher's (or the system's) intentions, role and actions; before, during and consequent to the administration of the task.

Attempts to model mathematical activity (both practice and thinking) in assessment systems led to the development of several multiple-component systems, such as the 12th grade mathematics assessment in use in Victoria, Australia, in the early 1990s, in which students completed a multiple-choice skills test, an extended answer analytic test, a 10-hour “Challenging Problem”, and a 20-hour “Investigative Project”. Each of these components was weighted equally and were intended in their totality to model mathematical activity in a variety of contexts and forms.

A danger exists that authentic assessment will join the other catch-phrases of recent times such as problem solving and constructivism and lose much of its power through simplistic and misguided misinterpretation. For example, to reduce authentic assessment to a requirement that students engage in “the mathematical activity that real folks do” (a description the author heard repeatedly in the USA throughout the 1990s) is to do a disservice to mathematics, to ignore the legitimately preliminary nature of schooling and to ignore the reality of students’ own mathematical activity. After all, students are real folks too.

An alternative view has been offered of “constructive assessment”:

> Constructive in the nature of the tasks such an approach would employ; constructive in the reconciliation of assessment with instruction, which gives authenticity to the assessment process; and constructive in the recognition that assessment must constructively inform the actions of the school community. (Clarke, 1992a, p.166).

Use of the term “constructive” in the context of assessment is more than a rhetorical device. Classrooms in which instructional practice is informed by constructivist learning theories are adopting the same imperative in the realization of their pedagogical goals as that which drives the constructive assessment initiative: the progressive transferral of the locus of authority from teacher to pupil. The enactment in practice of this shift in the locus of authority with respect to the control of and responsibility for student learning constitutes the most immediate challenge for classroom instruction and classroom assessment. Open-ended tasks, by their nature, offer students the opportunity to assume greater responsibility for the form of the response. The purpose of this chapter is to address the related questions as to whether open-ended tasks model mathematical thinking appropriately, offer students the opportunity to construct a personally-meaningful mathematics, reconcile instruction with assessment by providing learning contexts which instruct while offering insight into learning and whether these insights can usefully inform instructional practice and student learning.

If the modeling of mathematical thinking is the thesis behind recent assessment initiatives, then it should be acknowledged that much of the history of assessment in mathematics has resembled the antithesis of this goal.

3 The Assessment Agenda - Thesis or Antithesis?

> And the children in the apple-tree
> Not known, because not looked for
> But heard, half-heard in the stillness
> Between two waves of the sea.
> Quick, now, here, now, always. (Eliot, 1950, p.43)

Past assessment practices have been largely insensitive to the individuality of students' mathematical constructions. Assessments have offered brief glimpses of student performances and treated these glimpses as representative and even definitive. Assessment in mathematics has misrepresented both mathematics and the student's learning.

Assessment is fundamentally concerned with the exchange of information. If our assessment is to usefully empower teachers, pupils and the community in ways that will lead to more effective learning then the information exchanged in the assessment process must be quality information and it must be conveyed or documented in such a form as to usefully inform action.

Assessment in mathematics should:

1. Explicitly recognize all valued mathematical activities;
2. Model effective instruction and assessment;
3. Inform consequent action.

The notion that assessment practices, and particularly state-wide mandated practices, should serve the purpose of modelling good instructional and assessment practices arises from a vision of authentic assessment and a desire to model and develop "mathematical power" (California Department of Education, 1991). A further motivation comes from the recognition of the power of mandated assessment to act as a catalyst for change in educational systems (Barnes, Clarke, and Stephens, 2000). The use of the term "catalyst" is problematic in this context, where changes to mandatory assessment practices have coercive features that belie the neutrality suggested by the catalyst metaphor. The essential point is that our assessment should model the mathematical performances that we value.

The simultaneous modeling in the same activity of both good instruction and good assessment is consistent with the integrative aspect of authentic assessment. This modeling was particularly evident in the "Investigations" developed as part of the California Assessment Program (CAP), in which a preliminary pre-task activity provided students with an important familiarization with context and materials, while also offering an explicit exemplar of good instructional practice (see task type #6, later in this chapter).

While systemic change in assessment can legitimize new content and new practices, the further removed the assessing body is from the classroom the less likely it is that the assessment will usefully inform local practice. Accepting that the fundamental purpose of school assessment is to facilitate student learning, it is essential that assessment practices be structured to maximize the exchange of useful information leading to actions likely to achieve this purpose. The question posed in the California Mathematics Framework – "How can we tell whether student work demonstrates mathematical power?" - maintains the retrospective, summative emphasis, which has characterised mathematics assessment previously. Statements of assessment policy must address directly the need for assessment to inform teaching and learning. The New Standards Project (USA) explicitly endorsed such a proactive goal: "We were not out to measure performance but to greatly improve it" (1991, p. 7). However, the mechanism linking the intended assessment to any consequent actions that might improve teaching or learning outcomes is never described.

The slogan "No-one ever got any taller just through being measured" has become proverbial in education and variously attributed. It remains true, however, that the impact of curricular innovation on mathematics instruction and student learning will be minimal unless the obligation and the means for consequent action are explicit within the assessment system.

If our assessment is to usefully empower teachers, pupils and the community in ways that will lead to more effective learning, then the information exchanged in the assessment process must be quality information and it must be conveyed or documented in such a form as to usefully inform action. The means by which assessment information might be documented and exchanged have been variously detailed (CAP, 1991; Clarke, 1989; Clarke, 1992b; Stenmark, 1992). The criteria regarding the documentation and exchange of assessment information include:

1. The assessment must accurately portray the student's learning and, where possible, the experiences on which that learning was founded;
2. The assessment should indicate directions for future action by the teacher, pupil, parent or support person, which are likely to facilitate further learning.

Among the concerns related to criterion 1 is the capacity of the assessment process to recognize a range of learning styles and to be fair to all groups of students and free from bias. Similarly, any attempt to apply criterion 2 must recognize differences in various communities' capacity to provide support, the diverse-and extensive demands on a teacher's time, and the personal characteristics of the student, which may restrict their ability to act on the information provided. Both sets of concerns strengthen the argument for locating the entire assessment process as close to the classroom as possible. The question is: Do open-ended tasks offer the means to meet the above assessment criteria?

4 Task Alternatives

> There are three conditions which often look alike
> Yet differ completely, flourish in the same hedgerow:
> Attachment to self and to things and to persons, detachment
> From self and from things and from persons; and, growing
> between them, indifference
> Which resembles the others as death resembles life,
> Being between two lives - unflowering, between
> The live and the dead nettle. This is the use of memory:
> For liberation. (Eliot, 1950, p.40)

Contemporary mathematics instruction and contemporary assessment of mathematics learning must recognize and locate the learner as an individual, in relation to things mathematical, and in relation to the community in which learning occurs. The assessment of a student's learning involves the construction of a model of that learning.

> A model of individual behaviour must refer to more than just the actions, thoughts and beliefs of a single student, since those actions may only derive their meaning from their contribution to the realization of the group's goals; the thoughts lose significance if considered in isolation from the thoughts, motives and expectations of others; and the beliefs lose coherence once considered outside the societal context which gave them shape. (Clarke, 1992c, p.5)

The abstract constructs of mathematics may appear detached from their origins in human endeavour, however the teacher cannot afford either detachment or indifference, and must grasp the live nettle in recognition that it is the student's learning which is to be modeled, not the merely the mathematical knowledge which appears as a product of that learning. The employment of memory in the completion of mathematical tasks should be directed towards the goal of empowerment (or liberation), rather than recall or mimicry. Tasks are required which offer insight to both pupil and teacher, which facilitate the student's modeling of mathematics and the teacher's modeling of the student's learning.

It has been argued that existing tests in mathematics measure separate skills and concepts rather than the "knowledge and use of the interrelationships among mathematical ideas" (Webb and Romberg, 1988). In 1989, the Mathematical Sciences Education Board (MSEB) of the United States of America suggested that current tests "reinforce in students, teachers and the public the narrow image of mathematics as a subject with unique correct answers" (MSEB, 1989, p. 68).

Since then, large amounts of money and personnel have been committed to the development of tasks that meet the requirements of the newly-conceived assessment agenda. It is important to identify the range of task-types already available. I have set out one such listing below, in which examples of task prototypes are provided in a notional order of increasing complexity. It is noteworthy that many of these task types have been available to us for the past twenty years.

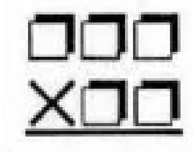
4.1 *Types of tasks*

Multiple-choice questions
(from Victorian Curriculum Assessment Board (VCAB), Reasoning and Data, Common Assessment Task 3, Facts and skills test, 1991)

Three students each have a probability of 0.8 of getting their assignments completed by the due date, independently of each other. The probability that none of the three assignments is completed by the due date is

A. 0.008 B. 0.2 C. 0.512 D. 0.6 E. 0.8

Enhanced multiple-choice questions
(from California Assessment Program, A *sampler of mathematics assessment, 1991)*



The five digits - 1,2,3,4, and 5 - are placed in the boxes above to form a multiplication problem. If they are placed to give a maximum product, the product will fall between:

A. 10,000 and 22,000
B. 22,001 and 22,300
C. 22,301 and 22,400
D. 22,401 and 22, 500

Numerical response question
(from Alberta Education, grade 12 diploma examination, 1992)

The first three terms of an arithmetic sequence are 19 - x, 3x, and 4x - 1. Correct to the nearest tenth, the numerical value of second term of this sequence is ____

"Good Questions" (open-ended, specific content domain)
(from Sullivan and Clarke, *Communication in the classroom: the importance of good questioning,* 1991a; see also *Language of Functions and Graphs,* Shell Centre, 1986)

The average of five numbers is 17.2; what might the numbers be?

The particular characteristics of this type of task are the content-specific focus, and the opportunity for answers at different levels of sophistication. In addition to mathematical correctness, student performance can take one of the following forms: the provision of only a single answer, the recognition of the existence of multiple answers; a comprehensive listing of all such answers, or the provision of a general statement encompassing all possible answers.

Extended answer question (explicit cueing and guidance)

(from VCAB, Space and Number, Common Assessment Task 4, Analysis task, 1991)

In 1989, the cost of a 'Culturpac' seven-day holiday package in Europe was determined by the distance of the destination from London.

Some examples are:

Destination	Q	R	S	T
distance (d)	625	2500	2916	1444
cost in $ (C)	700	1200	1280	960
$x = \sqrt{d}$	25	50	54	38

a. Plot all the ordered pairs (x, C) using the axes provided on the answer sheet and join the points. (Clearly label each point). Extend the line in both directions, for $0<x<60$.

b. If the equation of the line obtained is $C = mx + k$, find the values of m and k.

c. Write down the formula for C in terms of d and your values for k and m.

d. As of January 1st, 1990, all the prices were increased by 10 per cent.

i) What is the new formula connecting the cost with the distance?

An opposition tour operator, Sightseer, decides that for their 1990 tours the cost will be determined by the formula $C = 25x + 100$.

ii) Find the values of d for which Sightseer offers cheaper fares than Culturpac.

Open-ended, extended answer question (some cueing of method)
(from CAP, 1991 Test problem H-8, see also *A question of thinking, 1989*)

> A school mathematics club was designing games for students to play at a school carnival booth. One of the games was:
> Take two ordinary dice of different colours (for example, one white die and one red die). Roll both dice together. The student player wins if the number on the white die is greater than the number on the red die. The math club wins otherwise.
> Explain how you could decide whether or not [he players and the club have an equal chance of winning. Use a diagram if it helps clarify your explanation.

VCAB challenging problem (Problem solving, significant mathematical content)
(from VCAB, Change and Approximation, Common Assessment Task 2, Challenging Problem, 1990b)

> For any triangle there exists a point X such that the sum of the distances from each vertex to X is a minimum. Consider all isosceles triangles of perimeter 6 units. For which of these triangles is the sum of the distances from each vertex to X a minimum?

Fermi problems (context-specific, minimal cueing of mathematical tool skills)
(from the [Australian] Mathematics Curriculum and Teaching Program, 1988)

> How many piano tuners are there in Singapore?

It is the intention with this type of task that students should not have recourse to other sources of information beyond the knowledge of the group with which they are working. For the purposes of recording their problem solving attempts, the students were asked to employ a five-part report format:

a. State the problem
b. What do you know that would help you solve the problem?
c. Record what you did, step by step.
d. State your answer.
e. How good is your answer?

One of the characteristics which distinguishes this task-type is its capacity to reveal the mathematical tool skills which a student spontaneously chooses to access. This characteristic is shared by such problems as "Which is the better fit: a square peg in a round hole or a round peg in a square hole?" (from Schoenfeld, 1985). As with the Good Questions of task type #4, these tasks are characterized by simple expression, leaving the responsibility for the elaboration of the task demands in the hands of the student.

4.2 *Task selection*

In choosing a task-type for the purposes of assessment it is useful to subject the task to the scrutiny of these four questions:

a. What aspects of mathematical performance are being assessed by tasks of this type?
b. What elements of this task-type are essential to the purpose?
c. What elements of this task-type are optional for the purpose for which it is to be used?
d. What elements are missing from this task-type which might contribute to the purpose?

The preceding eight task types offer the opportunity for students to display a range of mathematical performances. The introduction of an assessment system that embodied the three principles given earlier would possibly employ a subset of the task-types outlined to model, in combination, the form of mathematical behaviour, competence, or power which the curriculum sought to promote.

Among those tasks illustrated above, categories 4, 6 and 8 are categorized as "open-ended tasks" for the purposes of this chapter. Any such categorization is open to challenge and examples provided are intended to be illustrative rather than definitive.

5 Open-ended Tasks - The Name of The Rose?

> We shall not cease from exploration
> And the end of all our exploring
> Will be to arrive where we started
> And know the place for the first time. (Eliot, 1950, p. 43)

What are the particular virtues of open-ended tasks? Consider the lines above as a poetic depiction of the problem solving process in mathematics and the insight that such activity affords. Do open-ended tasks offer such opportunities for exploration and insight? A necessary precursor to this question is the question, "What is an open-ended task?" The set of sample task types in the preceding section provides a more useful explication of the nature of open-ended tasks than a bald definition. The difficulty in defining open-ended tasks lies in the problematic nature of the context in which the task is framed. A task may allow for multiple answers, as is the case with Good Questions (#4 above), however if one answer is explicitly accorded higher value than another, then the task will be interpreted by the respondent as a closed task. Alternately, if the "open-ended" character of the task resides in the possibility of multiple solution pathways, then almost any task can be considered open-ended, since pupils' idiosyncratic solution techniques have been documented widely. A particularly notable example of this last phenomenon can be found in the paper "Two hundred ways to subtract: Most of them wrong" (Marriott, 1976), an analysis of 2826 children's attempts to subtract 586 from 940. Difficulties in the definition of open-ended tasks threaten to divert attention from the underlying rationale: the legitimate modeling of sophisticated mathematical activity in classrooms.

"Problem solving" has been plagued by similar semantic difficulties. Attempts at definition have communicated intention but hardly constitute unambiguous and succinct specification.

> For any student, a mathematical problem is a task
> (a) in which the student is interested and engaged and for which he wishes to obtain a resolution, and
> (b) for which the student does not have a readily accessible mathematical means by which to achieve that resolution.
> (Schoenfeld, 1989, pp.87-88)

Debate over the meaning of "readily accessible" and uncertainty regarding the origins of the student's affective response to the task (what if the student's motivation is "to get a better grade"?) distract from the success of this definition in invoking engagement and an implicit notion of non-routine tasks.

Clearly, a mathematical "Problem" need not be "open-ended". However, advocates of the use of problems and open-ended tasks draw on similar arguments, and a research-based justification of the use of open-ended tasks can call legitimately upon existing research into problem solving and problem-based curricula.

5.1 *Student responses to open-ended tasks - grasping the nettle*

> Through the unknown, remembered gate (Eliot, 1950, p.43)

In this chapter, assessment is portrayed as the process of modeling student learning through the observation of student mathematical activity in response to specific tasks. The paradoxical nature of many open-ended and problem solving tasks is that they require students to apply familiar and rehearsed skills in unfamiliar, non-routine contexts. It must be established whether student responses to open-ended mathematical tasks do provide accurate and appropriate information from which a model of the student's learning might be constructed. For the purposes of this discussion, one type of open-ended task will be taken as representative.

Use of the "Good Questions" task-type for both instruction and assessment has been investigated extensively (Clarke and Sullivan, 1992;

Sullivan and Clarke, 1988). These tasks have a particular character:

The average of five numbers is 17.2, what might the numbers be? A number is rounded off to 5.8, what might the number be? A rectangle has a perimeter of 30 units, what might be its area?

The characteristics of these Good Questions have been discussed elsewhere (Sullivan and Clarke, 1988; Clarke and Sullivan, 1990, Sullivan and Clarke, 1991a and b, Sullivan, Clarke, and Wallbridge, 1991). Each of the following postulated characteristics of Good Questions derives from a specific educational stance:

1. The task should require more than the recall or replication of a fact or procedure.
2. The task should be educative.
3. The task should be open-ended.

In studies of student responses to tasks such as "A number is rounded off to 5.8, what might the number be?" the following response coding was used:

0 = no response or an incorrect response;
1 = a single correct answer (for example, 5.79);
2 = several correct answers, without an obvious attempt to provide a systematic listing of possible answers (for example, 5.81, 5.83, 5.76);
3 = a systematic listing of possible correct answers (for example, 5.75, 5.76, 5.77, 5.78, 5.79, 5.80, 5.81, 5.82, 5.83, 5.84);
4 = a correct general statement specifying all possible answers (for example, "Any number equal to or bigger than 5.75 and smaller than 5.85").

Perhaps the most consistent and significant finding to emerge from research into the use of such "Good Questions" was the reluctance of pupils from grades 6 to 10 to provide more than a single answer to such open-ended questions. The term “reluctance” is used here to indicate the disinclination of the pupils tested to give either multiple answers or a general statement. This disinclination should not be confused with an inability to provide multiple answers. If open-ended tasks like those above were reworded so as to require pupils to list all possible multiple answers, then the number of multiple responses increased substantially. This demonstration of a capacity to provide multiple answers was also evident in the responses of pupils when the tasks were administered in interview situations. It is clear from these earlier studies (Clarke and Sullivan, 1992) that, while few students are equipped to produce general statements as solutions to open-ended tasks, many pupils have the capability to produce multiple answers but do not choose to do so.

Key findings of the research (from Sullivan, Clarke, and Wallbridge, 1991; Clarke and Sullivan, 1992) included:

1. The capability to give multiple and general responses increased with age.
2. Revision of the “Good Question” format to specifically request multiple answers resulted in a significant increase in the proportion of students providing multiple responses. However, even using this format there were many pupils who gave zero or one response.
3. Students were more likely to provide multiple responses in an interview context than in a test context.
4. Student responses to such tasks administered in test-like conditions were not significantly different after a seven-lesson teaching program based solely on the use of such open-ended mathematics tasks to the responses of students from a control class instructed to follow the program presented in the most commonly used text. Despite the explicit valuing of multiple responses in the instructional treatment, when responding to a test involving open-ended tasks that did not request multiple answers explicitly, there were no students in either group who attempted multiple answers.

The use of open-ended tasks to assess student learning and as instructional tools has been discussed and studied from the perspective of many academic disciplines (for instance, Thomas, 1989b. p. 44). With regard to the possibility of disadvantage, evidence exists that girls may be disadvantaged by the use of open-ended items in the context of science.

> We see once more that the open-ended nature of the question may have disconcerted many students .... Many fewer girls than boys attempted the question. (Thomas, 1989a, p. 43)

Research undertaken some time ago (Clarke, Sullivan, and Spandel, 1992; Sullivan, Clarke, Spandel, and Wallbridge, 1992) sought to investigate student responses to open-ended tasks with respect to the effects of grade level, gender, task-specificity, the presence or absence of explicit cueing, and the interaction of these variables with academic domain. In contrast with the findings of Thomas (1989a), girls in this study were significantly more likely than boys to provide multiple answers when specifically requested to do so, with the exception of mathematics in which the results were comparable.

Conclusions which might be drawn from these results included:

- that the inclination to give single responses (or the reluctance to give multiple responses) appeared to be a product of schooling, and not peculiar to mathematics. Both Year 7 and year 10 pupils were similarly reluctant to give multiple answers in all four academic contexts (mathematics, science, social studies and English);
- that the explicit request of multiple responses produced a significant increase in the quality of response (response level) in all four subjects;
- that the capability to give multiple responses increased significantly with year level, except in the context of English;
- that gender-related differences in response level were evident, and where these existed they favoured girls.

Several issues have been raised regarding the use of open-ended items for assessment purposes. In part, the motivation for their use in instruction can be aligned with the view of Cornbleth (1985).

> The myth of the right answer ... fosters an illusion of knowing that subverts construction of a functional knowledge base for critical thinking. (Cornbleth, 1985, p.31)

Within the mathematics education community, mathematical power has been identified with the capacity to solve non-routine problems (e.g., NCTM, 1989), and open-ended tasks are seen as an appropriate vehicle for instruction and assessment of students' learning in this regard. Given the various curriculum initiatives that have employed open-ended tasks for assessment purposes (for example, CAP, 1989; VCAB, 1990a), the results of this research assume some significance.

Arguments have been framed elsewhere in the literature, which assert that for the purposes of assessment the conclusions drawn from closed and open-ended items are sufficiently similar as to render the use of open-ended items unnecessary and undesirable on practical grounds.

> Whether students are tested via multiple-choice items or open-ended ones, their relative performance does not differ significantly.
> (Badger, 1990, p. 5)

This argument restricts the purposes of assessment to the ranking of students. However, present notions of assessment demand that assessment perform a variety of functions, the most central of which is the provision of information to inform the actions of learners and teachers. I would like to argue, in addition, that assessment based solely on closed or multiple-choice items will be insensitive to the process outcomes that have constituted much of contemporary mathematics curricula for the past 20 years (e.g., NCTM, 1989) and which are now being embraced even more widely in the mathematics curricula of several Asian countries.

Among the various concerns expressed about the use of open-ended tasks for assessment in mathematics, student unfamiliarity with the required mode of response is frequently cited. In response to this observation, it should be noted that, for the case of mathematics, even when multiple responses were given the explicit endorsement of an instructional intervention (seven lessons), no corresponding increase in response level ensued (Sullivan, Clarke, and Wallbridge, 1991). This would suggest that student reluctance or incapacity to provide multiple responses to open-ended tasks was attributable to more than simply lack of familiarity with the task format. It is possible that while the brief instructional intervention effectively addressed the issue of familiarity, student inclination to give single responses to academic questions is a consequence of an extensive educational history, and that a much more extended instructional intervention would be required to challenge this "training" effect. This study has demonstrated that this training effect is evident in other academic contexts besides mathematics, and may be more appropriately seen as a consequence of schooling rather than just mathematics instruction.

The legitimacy of relating student responses to non-routine and open-ended tasks to curricular content currently being studied continues to be the subject of research. It may not be realistic to expect students of any age to access recently-acquired skills in open-ended or problem-solving situations.

5.2 *Grading responses to open-ended tasks*

> A condition of complete simplicity
> (Costing not less than everything) (Eliot, 1950, p.44)

Holistic Assessment

Together with the acceptance of the inherently complex nature of mathematical activity and the need to provide classroom experience with "complex, thought-provoking" activities (New Standards Project, 1991) has come "holistic assessment" and the associated discarding of analytical scoring techniques. Advocates of holistic assessment sensibly argue that the interrelatedness of mathematical concepts and skills in the completion of a complex mathematical activity is best assessed holistically: that is, without the explicit attachment of numeric weights to solution components. Unfortunately, the abandonment of analytical scoring protocols has been accompanied in some instances by the abandonment of any analytical framework whereby the elements of a student's solution might be given explicit consideration. That is, the principle of holistic assessment has become confused with a commitment to non-analytic assessment. Non-analytic scoring has become synonymous with holistic assessment and has led also to the conviction that the assessment of a student's response to a complex mathematical task should not be documented in any form which employs an analytical framework such as the four dimensions of problem solving behaviour found in Schoenfeld (1985) and Clarke (1989), or the response coding employed for the study of Good Questions in this chapter. This stance is both naive and counter-productive. Clearly, the assessment of an essay or a work of art historically has been holistic in character, and yet in both cases a theoretical framework guides the assessment and facilitates the identification of excellence or mediocrity, without incurring any obligation to quantify elements within the analytical framework for the purposes of assigning a grade. Importantly, in the context of the classroom, the communication of an assessment through a single grade has little chance of constructively informing the future practices of either student or teacher and, lacking explanatory detail, runs the risk of appearing arbitrary.

If an obligation exists to record assessments in the form of grades, then the grade should be supplemented by a descriptive statement commensurate in its detail with the quality and detail of the information provided by the assessment tasks. It should also be made clear how the grade itself is to be interpreted.

With regard to the meaning that might be attached to the grade score, at least three approaches might be employed in answering the question, "What do you know if a student is a 4?"

1. Central tendency

"A student who is a 4 will most likely do these things" (examples of typical behaviours provided).

2. Exclusion
"A student who is a 4 will be able to do these things and will not be able to do these other things."

3. Range
"The performance of a student who is a 4 will lie within these performance bounds."

One must always ask, of course, whether it is the student who is a "4", or rather the student's response to a task or battery of tasks. Teachers will require support if they are to employ and/or interpret such grades. Based on Australian experience, the most useful form such support could take would be the provision of annotated student work samples.

The obligation imposed by state departments or school administrations to reduce assessment information to a single grade for the purposes of comparison or reporting is one reason why assessment has been practised typically as distinct from instruction, since the information which assessment has the potential to provide has been discarded in the grading process.

> To summarize assessment information... with a letter grade is to sacrifice precisely that detail which might most constructively contribute to the subsequent actions of teacher, student and parent (Clarke, 1989, p.5).

In summary, if any assessing body is committed to graded assessment then an obligation exists to articulate and illustrate the meaning of each grade assigned and to supplement a student's grade with descriptive detail that will usefully inform the consequent actions of teacher and student. If the assessing body, for whatever reason, were to restrict itself in its communication to schools and students to a single grade as the total measure and message regarding each student's performance in mathematics, the message would inevitably be a counterproductive one. Where the assessment task is open-ended, any such simplification buys simplicity at the cost of almost all the informative detail that characterizes student responses to open-ended tasks.

6 Conclusions and Synthesis

Even if we were to concede the instructional virtues of the use of open-ended tasks, their use for assessment purposes in mathematics requires justification. The summation which follows sets out some of the points for and against the use of open-ended tasks for mathematics assessment. Each point highlights an issue central to the use of open-ended tasks for assessment. Careful consideration of each point would inform the design and use of open-ended tasks for assessment purposes.

6.1 *Why use open-ended tasks/or assessment purposes in mathematics?*

Answer 1. Legitimate mathematical activity
Because such tasks model mathematical activity appropriately and effectively;

Answer 2. Acting as a catalyst for systemic curricular change
Because such tasks, once endorsed by a state or national assessment body, act as a catalyst for systemic change of the official mathematics curriculum and of the taught curriculum within mathematics classrooms;

Answer 3. Stimulating teacher professional growth
Because the legitimization of such tasks through their inclusion in assessment practices necessitates teacher classroom experimentation with associated instructional practices, and increased the likelihood of teacher professional growth (Clarke and Hollingsworth, 2002; Clarke and Peter, 1993);

Answer 4. The individuality of student learning
Because such tasks offer students the opportunity to construct responses in ways that reflect the individuality of each student's mathematical understandings and so offer insight into student learning;

Answer 5. The complexity of mathematical thinking
Because such tasks require complex responses at a level commensurate with the complexity of the mathematical thinking we seek to monitor: both elaborated responses over extended time periods, and sophisticated responses synthesizing a range of concepts and procedures;

Answer 6. Spontaneous skills access
Because such tasks are more likely than conventional closed tasks to indicate the mathematical tool skills that a student chooses to access, and thereby provide a more accurate indication of the developing mathematical power of the student than would be the case with specifically-cued skills items;

Answer 7. Detailed information
Because student responses to such tasks provide information in sufficient detail to meaningfully inform consequent instructional practice;
Answer 8. Student engagement

Because students find the contextual location and the challenge of most open-ended tasks more stimulating and interesting than conventional closed tasks, with a consequent increase in student engagement with such tasks.

6.2 *Why avoid using open-ended tasks for assessment purposes in mathematics?*

Answer 1. Misrepresent the student through a minimalist response
Because student responses to such tasks may misrepresent the student's mathematical capabilities, where the student supplies a correct but minimalist response;

Answer 2. Misrepresent the student through contextual difficulty
Because open-ended tasks are more likely to have elaborate contexts, student familiarity with the specific context of a task is a more significant factor in determining the quality of a student's response than is the case with conventional closed tasks;

Answer 3. Misrepresent the student through difficulties of literacy
Because open-ended tasks typically take more elaborate verbal forms than closed tasks, the mathematical understandings of students with language difficulties will be misrepresented in their responses to such open-ended tasks;

Answer 4. Misrepresent the student through inadequate behaviour sampling
Because open-ended tasks typically require longer periods of completion time, less tasks may be used for the purposes of student assessment, which is consequently based upon a narrower behaviour sample, and more likely to be influenced by task-respondent interactive effects (such as context, as noted above);

Answer 5. Questionable legitimacy of grading
Because student responses to such tasks are typically complex to a degree that precludes the simplistic quantitative grading required by school systems: that is, questionable grading validity;

Answer 6. Complexity of information
Because student responses display a complexity which defies useful summation for the purposes of informing instructional action;

Answer 7. Lack of reliability
Because student responses display a complexity and a degree of idiosyncrasy that makes problematic the reliability of any assessment scoring;

Answer 8. Relative performance is no different
Because whether students are tested via closed or open-ended tasks, their relative performance does not differ significantly;

Answer 9. Lack of teacher expertise
Because teachers lack experience in the use of such tasks, substantial in-service and pre-service programs will be required if teachers are to use such tasks appropriately and effectively;

Answer 10. Student stress

Because students find open-ended tasks more demanding and consequently more threatening and stressful than conventional closed items, where expectations and methods are specified in more detail.

6.3 *The effective integration of open-ended tasks into assessment practices*

The three principles stated earlier in this chapter, constitute one attempt to provide simple criteria for the construction of an assessment system that will adequately represent both mathematics and the student in a way likely to constructively inform teaching and learning. The role of open-ended tasks within such a system will depend upon our success in addressing the concerns listed above.

In concluding, I would like to offer some practical suggestions for future directions in which we might usefully expend some effort. I have argued that a substantial range of prototype mathematical tasks now exists sufficient to meet our assessment needs. These are the tools of sophisticated mathematical assessment. But these tools have been in existence for twenty years. Why do we not see more widespread use of open-ended tasks in mathematics assessment? It could be that the necessary systemic approach has yet to be developed. This is the challenge for the present generation of educators.

First, we require straightforward criteria for the selection of suitable batteries of such tasks. Such criteria should include the requirement that the selected tasks:

- reveal tool skill acquisition and deployment;
- provide the opportunity for sophisticated performance;
- provide immediate feedback to students and teachers;
- constitute in combination an evolving profile of student mathematical performance;
- constitute effective instruction;
- provide easily documented information.

I would suggest that various specific types of open-ended tasks in mathematics have the potential to meet these criteria. An effective first step is the development of a large bank of such tasks containing sample items suitable for all topics and grade levels.

Second, we require a school community expert in the use of such tools/tasks. Both teachers and students will require extensive practice and support in the use of such tasks. both in instruction and assessment contexts (see Clarke, 1995). Part of this process will involve the recognition that the same tasks can serve both instructional and assessment purposes, and summative and formative purposes. The reconceptualisation of formative assessment as central to effective instruction is presently underway in many school systems around the world. The documentation of student performance should serve each and all purposes.

Third, we require an assessment regime whereby the information generated by such tasks can be put to good use. If assessment is to play the constructive role within the curriculum that is the stated goal of recent assessment initiatives, optimal communication mechanisms will be required to inform (in the short-term) the actions of students, teachers and parents, and (in the longer term) school administrations, curriculum developers, and State boards of education.

6.4 *A social contract*

It should be possible to negotiate a social contract with teachers in which new modes of assessment are seen to offer teachers, schools, and education systems a mutually beneficial relationship. Under such a contract, the assessing body undertakes to provide quality assessment instruments (and, implicitly, quality instructional models), together with guidelines and support for teachers' cumulative monitoring of student performance. In return, teachers maintain the program of on-going monitoring, provide the authority with descriptive statements to supplement graded assessments, and administer the assessment instruments. The scoring of student performance on the new instruments is best undertaken by classroom teachers under the guidance, co-ordination and support of the assessing body. The individual student report arising from such an assessment system should combine informed scoring with teacher descriptive reports in a constructive portrayal of the student's mathematical power. Given this information, effective summative reporting will be possible at the level of the school, the district and the State.

If such a system were introduced, it is likely that one consequence would be that teachers at all grade levels would model their instruction and assessment on the new tasks. We must be convinced that open-ended tasks:

- model mathematical behaviour appropriately;
- constitute effective instruction and assessment;
- inform subsequent action.

If researchers and those involved in the development of assessment tools and systems can successfully address the concerns that have been raised in this chapter, then open-ended tasks may lead to the reconciliation of instruction and assessment, and provide the means by which sophisticated mathematical thinking is engaged in universally and routinely in mathematics classrooms at all levels of schooling: That is, a comprehensive and complete synthesis of our educational agendas.

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.
Through the unknown, remembered gate
When the last of earth left to discover
Is that which was the beginning;
At the source of the longest river
The voice of the hidden waterfall
And the children in the apple-tree
Not known, because not looked for
But heard, half-heard, in the stillness
Between two waves of the sea.
Quick. now, here. now, always -

A condition of complete simplicity
(Costing not less than everything)
And all shall be well and
All manner of thing shall be well
When the tongues of flame are in-folded
Into the crowned knot of fire
And the fire and the rose are one. (Eliot, 1950, pp. 43-44)

References

Badger. E. (1990, April). *Using different spectacles to look at student achievement*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA (USA).

Barnes, M., Clarke, D. J. & Stephens, W. M. (2000). Assessment as the engine of systemic reform. *Journal of Curriculum Studies 32*(5), 623-650.

Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan 80*(2), 139-144, 146-148.

Californian Assessment Program (CAP). (1989). A *question of thinking*. Sacramento, CA: California State Department of Education.

California Assessment Program (CAP). (1991). *A sampler of mathematics assessment.* Sacramento, CA: California Department of Education.

California Department of Education. (1991). *California mathematics framework.* Sacramento, CA: California Department of Education.

Clarke, D. J. (1989). *Assessment alternatives in mathematics.* Carlton, Vic.: Curriculum Corporation.

Clarke. D. J. (1992a). The role of assessment in determining mathematics performance. In G. Leder (Ed.), *Assessment and learning of mathematics* (pp.145-168). Hawthorn: Australian Council for Educational Research.

Clarke, D. J. (l992b). Activating assessment alternatives in mathematics. *Arithmetic Teacher, 39*(6), 24-29.

Clarke, D. J. (1992c, March). *Finding structure in diversity: the study of mathematical behaviour*. Paper presented to the National Council of Teachers of Mathematics, Research Pre-session to the 70th Annual Meeting, Nashville, Tennessee (USA).

Clarke, D. J. (1995). Quality mathematics: How can we tell? *The Mathematics Teacher, 88*(4), 326-328.

Clarke, D. J. (1996). Assessment. In A. Bishop (Ed.), *International handbook of mathematics education* (pp. 327-370). Dordrecht, The Netherlands: Kluwer.

Clarke, D. J., Breed, M., & Fraser, S. (2004). The consequences of a problem-based mathematics curriculum. *The Mathematics Educator*, *14*(2), 7-16 (published through the University of Georgia-Athens, USA).

Clarke, D. J. & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947-967.

Clarke, D. J. & Peter, A. (1993, April). *Classroom experimentation and teacher reflection in a dynamic model of professional growth*. Paper presented at the 1993 Annual Conference of the American Educational Research Association, Atlanta, Georgia (USA).

Clarke, D. J. & Sullivan, P. (1990). Is a question the best answer? *The Australian Mathematics Teacher, 46*(3), 30-33.

Clarke, D. J. & Sullivan, P. A. (1992). The assessment implications of open-ended tasks in mathematics. In M. Stephens & J. Izard (Eds.), *Reshaping assessment practices: Assessment in the mathematical sciences under challenge* (pp.161-179). Hawthorn: Australian Council for Educational Research.

Clarke, D. J., Sullivan, P. A., & Spandel, U. (1992). *Student response characteristics to open-ended tasks in mathematical and other academic contexts.* Research Report No.7. Oakleigh, Vic.: MTLC.

Cornbleth, C. (1985). Teacher questioning and student interaction: An observation of three social studies classes. *The social studies, 77*(3), 119-122. Cited in Pollack (1988).

De Lange, J. (1987). *Mathematics, insight and meaning*. Utrecht: Freudenthal Institute.

Eliot, T. S. (1950). *Four quartets.* (all excerpts quoted are from “Little Gidding”). Glasgow: The University Press Glasgow.

Lovitt, C. & Clarke, D. M. (1988). *The mathematics curriculum and teaching program.* Canberra: Curriculum Corporation.

Mathematical Sciences Education Board (MSEB). (1989). *Everybody counts: A report to the nation on the future of mathematics education*. Washington, DC: National Academy Press.

Marriott, P. (1976). *Two hundred ways to subtract – most of them wrong*. Paper submitted as part of the course requirements of the Bachelor of Special Education. Clayton: Monash University.

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

New Standards Project. (1991). Proposal for funding submitted to the National Science Foundation.

Schoenfeld, A. (1985). *Mathematical problem solving.* Orlando, FL: Academic Press.

Schoenfeld, A. (1989). Teaching mathematical thinking and problem solving. In L. B. Resnick and L. E. Klopfer (Eds.), *Toward the thinking curriculum: Current cognitive research* (pp.83-103). 1989 Yearbook of the Association for Supervision and Curriculum Development.

Shell Centre. (1986). The *language of functions and graphs.* Nottingham, UK: Author.

Shepard, L. A. (1991). Psychometrician's beliefs about learning. *Educational Researcher*, *20*(6), 2-16.

Stenmark, J. K. (Ed.). (1992). *Mathematics assessment: Myths, models, good questions, and practical suggestions.* Reston, VA: Author.

Sullivan, P. & Clarke, D. J. (1988). Asking better questions. *Journal of Science and Mathematics Education in South East Asia,* June, 14-19.

Sullivan, P. & Clarke, D. J. (1991a). *Communication in the classroom: The importance of good questioning.* Geelong: Deakin University Press.

Sullivan, P. & Clarke, D. J. (1991b). Catering to all abilities through the use of "Good" questions. *Arithmetic Teacher, 39*(2), 14-21.

Sullivan, P., Clarke, D. J., Spandel, U., & Wallbridge, M. (1992). *Using content specific open questions as a basis of instruction: A classroom experiment. Research Report No.4.* Oakleigh: Mathematics Teaching and Learning Centre (MTLC), Australian Catholic University (Victoria).

Sullivan, P., Clarke, D. J., & Wallbridge, M. (1991). *Problem solving with conventional mathematics content: Responses of pupils to open mathematical tasks. Research Report* 1. Oakleigh: Mathematics Teaching and Learning Centre (MTLC), Australian Catholic University (Victoria).

Thomas, B. (1989a). *On their own—Student responses to open-ended tests in science*. Massachusetts State Department of Education, Quincy Bureau of Research, Planning and Evaluation.

Thomas, B. (1989b). *On their own—Student responses to open-ended tests in social studies*. Massachusetts State Department of Education, Quincy Bureau of Research, Planning and Evaluation.

Törner, G., Schoenfeld, A. H., & Reiss, K. M. (Eds.). (2007) Problem solving around the world: Summing up the state of the art. Special Issue of the journal *ZDM: The International Journal on Mathematics Education, 39*(5-6), 353-561.

Victorian Curriculum and Assessment Board (VCAB). (1990a) *Mathematics: Study design.* Melbourne: Author.

Victorian Curriculum and Assessment Board (VCAB). (1990b). *Various 12th grade "Common assessment tasks"*. Melbourne, Victoria: Author.

Victorian Curriculum and Assessment Board (VCAB). (1991). *Various 12th grade "Common assessment tasks"*. Melbourne, Victoria: Author.

Webb, N., & Romberg, T. (1988, April). *Implications of the NCTM standards for mathematics assessment*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana (USA).

Chapter 8

Using ICT to Improve Assessment

Marja van den HEUVEL-PANHUIZEN

Angeliki KOLOVOU   Marjolijn PELTENBURG

This chapter addresses how ICT can be used to improve the assessment of students' understanding of mathematics. In general, we see three ways in which ICT can contribute towards better assessment: (a) making it possible to use tasks with high mathematical demand, (b) making tasks more accessible for students, and (c) revealing students' thinking and solution processes. We illustrate these new ICT-generated opportunities for assessment by providing examples from our own research projects in the field of problem solving, i.e., early algebra in the upper grades of primary school, and from solving subtraction problems with crossing the ten by students in primary special education who are weak in mathematics.

1 Introduction

Assessment and education have a reciprocal relationship. On the one hand, assessment is determined by the curriculum and on the other hand, assessment has a strong effect on what is taught and how it is taught (Pellegrino, Chudowsky, and Glaser, 2001; Popham, 2000). The latter means that assessment has the potential to be a lever for raising the quality of mathematics education and to be a tool for systemic innovation (Van den Heuvel-Panhuizen and Becker, 2003). In other words, with better assessment, the disadvantage of "teaching to the test" can become an advantage (De Lange, 1992a). Therefore, working on the improvement of assessment actually means working on the improvement of education.

Following Popham (2000, p. 3), we consider assessment as "a process by which educators use students' responses to specially created or naturally occurring stimuli in order to make inferences about students' knowledge, skills, or affective status." These inferences can lead to decisions with respect to individual students, complete classes or schools, or to the educational system as a whole.

At all these levels, assessment can be used for different purposes. The main distinction in this regard is Scriven's (1967) differentiation of summative and formative assessment. Summative assessment refers to the assessment that occurs at the end of a school year, a particular course or program. Usually these assessments are meant for certification or accountability purposes. Another example of summative assessment is judging students' performance against national standards. Contrary to summative assessment, formative assessment is done during a program or at its beginning. This assessment primarily informs teachers about their students' learning process, though it can also inform the students themselves. Mostly, the results are intended to adapt the program to the students' needs. Classroom assessment is one of the most common types of formative assessment. Since the end of the 1990s, formative assessment is strongly connected to the term "*Assessment for learning*" (AfL) (Assessment Reform Group, 2002). The AfL movement was especially prompted by studies carried out by Black and Wiliam (1998a, 1998b) in which they showed that improvement in classroom assessment practice has a strong effect on student achievements.

Two aspects of assessment that are crucial in achieving a better assessment are its content and format (Van den Heuvel-Panhuizen, 2007; see also Shepard, 2000). The content refers to the tasks given to students. These tasks should be mathematically meaningful and offer students the opportunity to apply their mathematical knowledge. In fact, the mathematics task that is presented to students forms the heart of an assessment in mathematics education (De Lange, 1992b). The task embodies what we think is important for students to learn. Moreover, the task determines to a large degree the nature of the student's response and, consequently, whether the assessment gathers the information that teachers need for didactical decision making.

Besides the task, the format is also a determining factor. The assessment format influences heavily the types of tasks that can be offered to the students and the kind of responses that can be collected from them. For example, paper-and-pencil assessment focuses more on retrieving students' answers than on revealing their cognitive processes.

The two formats that traditionally have been used in assessment are oral assessment and written assessment. Currently, assessment is enriched with a third format: assessment based on Information and Communication Technology (ICT) (Van den Heuvel-Panhuizen, 2007). This format brings in a number of unique opportunities for improving assessment from the perspective of the teacher as well as from that of the student. These unique opportunities of ICT to improve the quality of assessment and, consequently, that of teaching and learning mathematics, are discussed in this chapter.

In recent years, the use of ICT has fundamentally changed the spectrum of assessment possibilities. While in the early stages of using computers for assessment, tests on the computer were rather similar to the written tests (Bennett, 1998) – where questions on paper were converted to questions on screen – the recent advances in the information technologies brought in new possibilities for assessment, which clearly go beyond the written format and constitute a new assessment environment (Burkhardt and Pead, 2003; Ripley, 2003). In other words, current ICT-based assessment uses technologies more optimally. In our view, this can be done in three ways, which respectively results in (a) using tasks that have a high mathematical demand, (b) making tasks more accessible for students, and (c) revealing students' thinking and solution processes.

In this chapter, we give examples from our own research projects to illustrate these new opportunities for improving assessment by using ICT. It is important to realize that these examples are not derived from classroom practice and may not be directly applicable in school. Nevertheless, the main reason for including them in a yearbook for teachers is to give them a broader view on what is possible in assessment, a view that goes beyond common ways of testing.

2 ICT Can Make it Possible for Assessment Tasks to Have High Mathematical Demand

Computer-based technologies have highly influenced how and what students learn in school (Roschelle, Pea, Hoadley, Gordin, and Means, 2000) and have also changed assessment. One of the major opportunities that ICT-based assessment provides is a rich and natural working environment for students to work on complex tasks. An ICT environment makes it possible to present tasks that cannot be presented to students through a conventional paper-and-pencil format. This is so as the computer can interactively carry out simulations in which different but related values are dynamically linked. These computer simulations of problem situations are considered to be a rich and highly promising genre of ICT-based assessment (Burkhardt and Pead, 2003) by which skills and processes that are difficult to document using traditional tests can be assessed (Pellegrino, Chudowsky, and Glaser, 2001).

The following sample task, taken from the POPO project[1], clearly demonstrates that an ICT environment opens up opportunities for more complex assessment tasks. The so-called "*n* hits, *n* misses, and *n* points" task and the corresponding digital environment including the game *Hit the target*[2] (see Figure 1a) is meant to investigate primary school students' ability to solve early algebra problems. More precisely, the task is developed to assess whether students can deal with covarying variables.

In the task, which is about an archery game, the students have to figure out which game rule should be applied to get 15 points in total with 15 arrows that hit the target and 15 arrows that missed it. After this, the students are asked whether other game rules are possible that give the same result. Later, similar questions are posed with respect to 16 points, 16 hits and 16 misses, and 100 points, 100 hits and 100 misses. In fact, these problems have an infinite set of solutions that can be described by

---

[1] POPO stands for *Problem Solving in Primary School*. This project is being carried out at the Freudenthal Institute. Angeliki Kolovou and Marja van den Heuvel-Panhuizen are the main investigators of this project.

[2] The game *Hit the target* is developed by Marja van den Heuvel-Panhuizen and programmed by our colleague Huub Nilwik at the Freudenthal Institute.

the general relationship between the quantities of the problem: to get *n* hits, *n* misses, and *n* points the sum of the points per hit and the points per miss should be 1.The idea behind the assessment is to examine whether students are able to discover the general principle of getting an equal number of hits, misses, and total points.

Students in primary school who have not yet been taught formal algebraic strategies are not likely to solve the problem of getting *n* hits, *n* misses, and *n* points in a straightforward way by setting up the equation $15a + 15b = 15$. Instead, they have to reason about the relations between the hits, the misses, the game rule, and the total score. The game *Hit the target* provides them with an environment to do so.

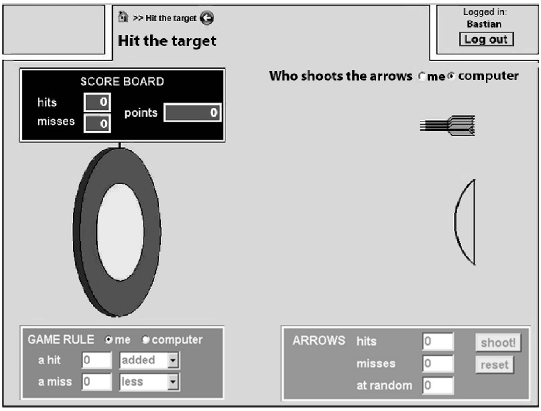


Figure 1a. Screen view of game in the computer shooting-mode

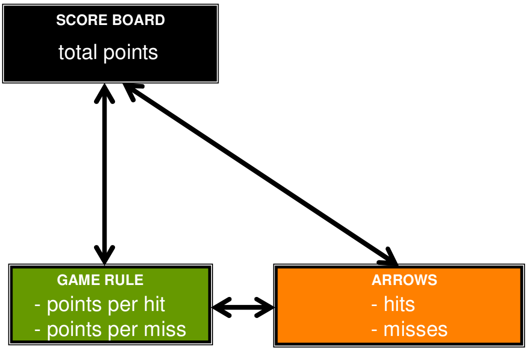


Figure 1b. Covarying variables

The screen of the game displays five features: the target, the pile of arrows and the bow, the score board that shows the points, the board that displays the game rule, and the board that displays the number of hits and misses. In the game the students can set the shooting mode and the game rule mode. The shooting mode has two options: shooting arrows one by one by dragging them to the bow (i.e., user shooting) or entering a number of hits and misses and shooting them at once (i.e., computer shooting). The game rule mode also has two options: the students can set the game rule by filling in the points added or subtracted in case of a hit or a miss (i.e., user defined game rule) or the computer sets the rule randomly (i.e., computer defined game rule). The number of arrows, the game rule and the number of points are dynamically linked. This means that a modification in the value of one of the variables has a direct effect on the other variables (see Figure 1b). While shooting or removing arrows from the target, the values on the scoreboard update rapidly in relation to the game rule. The students can use this instant visual feedback that shows the consequences of their actions as a means to perform self-assessment and error detection and as a source for constructing a model of the relationship between the variables (cf. Kolovou and Van den Heuvel-Panhuizen, 2010; Nathan, 1998).

The "*n* hits, *n* misses, and *n* points" example has shown that an ICT-based assessment can bring in more complex tasks that cannot be presented to students through a conventional paper-and-pencil format. In addition, we should always be aware that irrespective of the complexity of the problems the assessment format always influences the ability that is assessed (Bennett, Braswell, Oranje, Sandene, Kaplan, and Yan, 2008; Clariana and Wallace, 2002). For example, the interactive character of the computer environment which is lacking in a paper-and-pencil format may affect the student's performance and lead to different assessment results (Burkhardt and Pead, 2003).

3 ICT Can Make Tasks More Accessible to Students

ICT-based assessment can make problems more accessible to students, because it has the possibility to incorporate particular features that can help students to better understand what the tasks are about. In addition to static drawings and photographs that can be included in written assessment, ICT-based assessment can also contain dynamic video clips (Bottge, Rueda, Kwon, Grant, and LaRoque, 2007) and animations (Burkhardt and Pead, 2003).

Bottge et al. (2007) found that ICT-based assessment in which the problems were presented by means of a video reduced the cognitive load for low achieving students. The video-enriched assessment enabled the students to better demonstrate their understanding of the mathematical concepts they had been taught. Furthermore, struggling readers may benefit from using audio tools that read aloud text (Elbaum, 2007; Helwig, Rozek-Tedesco, and Tindal, 2002).

Special types of interactive, dynamic tools that can be added to an ICT-based assessment are mathematical auxiliary tools. In case students get stuck in solving tasks, they can mobilize these tools to carry out specific mathematical operations on the screen (Sarama and Clements, 2006).

3.1 *Mathematical auxiliary tools in ICT-based assessment*

In the following part of the chapter, we illustrate the use of mathematical auxiliary tools in an ICT-based assessment. The examples are taken from the IMPULSE project.[3] We discuss two types of tools: one including digital manipulatives and one consisting of a digital empty number line. Both tools are aimed at supporting students in solving subtractions problems up to 100 and, in particular, subtraction problems that require crossing the ten. These are problems in which the ones-digit of the subtrahend is larger than the ones-digit of the minuend (e.g., 48–39). A frequently made mistake in these problems is processing the ones-digits in the reverse way (e.g., in the case of 48–39, this means subtracting 8 from 9 instead of 9 from 8).

---

[3] IMPULSE stands for *Inquiring Mathematical Potential and Unexploited Learning of Special Education* students. This project is being carried out at the Freudenthal Institute. Marjolijn Peltenburg and Marja van den Heuvel-Panhuizen are the investigators of this project.

Another characteristic of these auxiliary tools, as used in our assessment, is that they are optional. Whether the students use them or not is their own decision. This means that the tools can stimulate the students to reflect on their own abilities to solve particular tasks and make them more active when being assessed. In fact, this can be considered as a form of self-assessment, like in the *Hit the target* game where the students have to interpret the consequences of their actions as shown on the scoreboard. With respect to the auxiliary tools, it is their optional character that brings in self-assessment in a natural way.

3.2 *Digital manipulatives*

Figure 2a shows what the students see on the screen of their computer when they have to solve the problem of the Spiroe comics (Janssen, Scheltens, and Kraemer, 2005). The calculation to be carried out is 48–39. If they want to use the auxiliary tool, they have to click the tool button. The digital manipulatives tool pops up (see Figure 2b). It consists of a digital 100-board and a storage tray with counters. Any number of counters from 1 to 10 can be dragged in one movement to the board. The board has a 10 by 10 grid, with a 5 by 5 structure, which means that the board is divided into four parts.

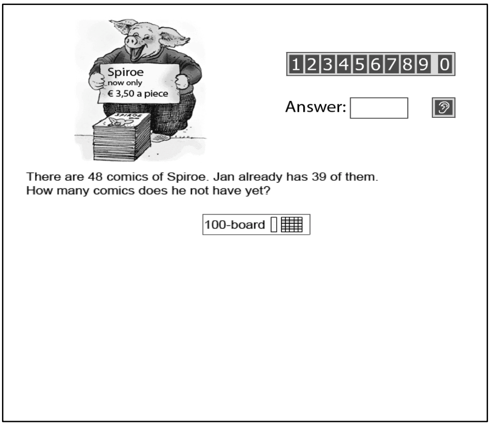


Figure 2a. ICT-based assessment with button for digital manipulatives tool

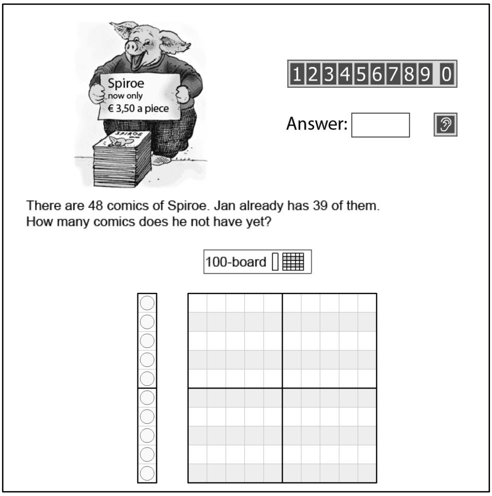


Figure 2b. ICT-based assessment with digital manipulatives tool popped up

The students can use the tool to make a visual representation of the numbers involved in a problem by putting the corresponding number of counters on the digital board and carrying out the required operations by moving or removing the counters. In contrast to working with wooden manipulatives on their desk, the digital counters provide flexible and manageable manipulatives, which might put students on the track of finding a solution by making use of the 5- and 10-structure instead of counting one-by-one.

Moreover, by creating an on-screen visual representation of the subtraction problem that has to be carried out, students may be less inclined to reverse the processing of the ones-digits. For example, in the case of 48–39, the tool can prompt them to find a solution for subtracting 9 from 8 by opening up the next ten. In this way, the auxiliary tool assists the students in modeling problems, which means that the tool provides them with a structure for thinking and acting when solving them (Bottino and Chiappini, 1998; Clements, 2002; Clements and McMillen, 1996).

3.3 *Digital empty number line*

The digital empty number line tool consists of a horizontal line, a pencil, an eraser, and a clear-all button (Figure 3). The tool operates by touch screen technology. The students can use the pencil to carry out operations by putting markers and numbers on the number line and making jumps backwards and forwards. This means that, as with the digital manipulatives tool, a flexible and easily manageable tool is offered to students. They can work with the tool in the same way as with a pencil on a piece of scrap paper.

Figure 3. Digital empty number line

The digital empty number line can help students in solving subtraction problems, since it can function as an aid to order the numbers involved in the problems and carry out the necessary operations. For example, in a problem in which students have to solve the subtraction problem 62–58, they have to think about the position of these numbers in the counting sequence, and then realize that these numbers are actually quite close to each other. This understanding can trigger them to bridge the difference by adding on from the subtrahend until the minuend is reached (see Figure 4).

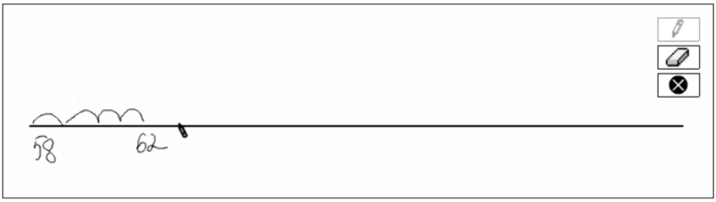


Figure 4. Solving 62–58 by adding on

Of course, students can also apply other strategies. For example, they can take away 60 (six times ten) from 62, arriving at 2, and then add 2 to reach the answer 4. This way of solving the problem is based on compensating.

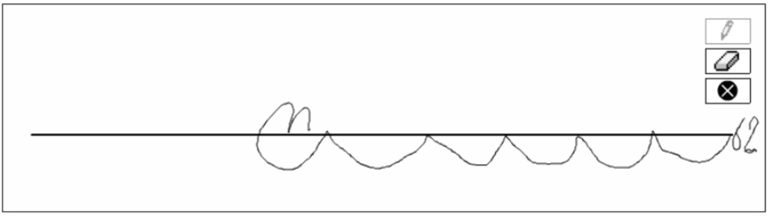


Figure 5. Solving 62–58 by compensating

3.4 *Mathematical auxiliary tools make tasks more accessible to students*

Our studies in which we assessed special education students' ability in solving subtraction problems with "crossing the ten" (see Peltenburg, Van den Heuvel-Panhuizen, and Robitzsch, 2010) verify that ICT can make tasks more accessible to students. The students in our studies were 8 to 12 years old and their mathematical level was one to four years behind the level of their peer group in regular education. We tested the students two times. First, they did a set of subtraction items in an ICT-based environment including an optional auxiliary tool (either the manipulatives tools or the digital empty number line tool)[4] and five weeks later they did the written standardized version of the test items, not including a tool.

A comparison of the students' responses from the two test formats showed that the ICT-based assessment was a more appropriate instrument than the standardized test to reveal what the students were really capable of. The students solved more problems correctly in the ICT-based assessment than in the standardized test. In addition, it

[4] The ICT-based environment including the optimal auxiliary tools is developed by Marja van den Heuvel-Panhuizen and Marjolijn Peltenburg; it is programmed by our colleague Barrie Kersbergen at the Freudenthal Institute.

appeared that the students were quite capable of judging their mathematical competence and therefore could decide whether the use of an auxiliary tool could be beneficial. This conclusion is based on the fact that the students who gave an incorrect answer to an item in the standardized test had used the tool more frequently for solving that item, than the students who gave a correct answer to that item in the standardized test format. In other words, when doing the firstly administered ICT-based assessment the students correctly judged their need for an auxiliary tool.

Based on these results, we could conclude that using ICT-based mathematical auxiliary tools like digital manipulatives and a digital empty number line can make tasks more accessible to students. These tools give students a better opportunity to show what they are able to do. Of course, solving these subtraction problems with the help of a tool does not reflect the same ability level as solving the problems without any help. So, one may wonder where this assessment with auxiliary tools is useful for. The power of this assessment is that it opens the students' "zone of proximal development" (Vygotsky, 1978) and that it better reveals what students are able to do. Moreover, whether the students used the auxiliary tools and how they used them informs us about what instruction is needed to bring students to a higher level of performance.

4 ICT Can Reveal Students' Thinking and Solution Processes

Although (video-recorded) observations and thinking-aloud procedures applied in oral assessment, offer good opportunities for gathering knowledge about the students' thinking and their way of solving tasks, these opportunities can be enhanced within ICT-based assessment. The use of computers makes it possible to follow and register the students' working very precisely, either by software that enables audio and screen recordings or by software that produces log files.

4.1 *Audio and screen recordings*

An example of the use of audio and screen recording software can be found in the Impulse project. Stills from the screen videos made of the students working in the assessment environment are shown in Figures 4 and 5. The screen videos allowed us to follow very precisely the students' movements on the computer screen when using the auxiliary tools for solving the subtraction problems. Another example that shows how informative screen videos are about the student's solution processes, can be found, for example, in Barmby, Harries, Higgins, and Suggate (2009). Their experiences show that these screen videos enable the examination of students' strategies in a more in-depth way than paper-and-pencil tests can do.

4.2 *Log files*

The use of software that produces log files, an approach that was applied in the POPO project, also make it possible to track and register students' workings. In order to get these log files, the online environment of the *Hit the target* game was linked to the Digital Mathematics Environment (DME)[5], which is a software that registers students' actions while working online. One of the log files from this project is displayed in Figure 6. It shows a student's working on the "*n* hits, *n* misses, and *n* points" task in the *Hit the target* environment. The advantage of this log-file approach, compared to making screen videos, is that having log files does not require real-time analysis of the student's working. These log files show the solution process in a more condensed way, but consequently they are less precise.

A log file consists of a list of all events that a student has carried out in the environment structured in sessions. A session contains the activity that a student performs each time she or he logs in the online environment. The date and the duration of the sessions are also registered. An event is a single shooting action performed when the

---

[5] The Digital Mathematics Environment (DME) is developed by our colleague Peter Boon at the Freudenthal Institute.

student clicks on the shoot button. By examining the succession of events that a student performs we can detect whether she or he tries to answer a particular problem or carries out random events. We call the events connected to answering a problem focused events.

In the log file in Figure 6 we can read that in event 1, the student chose the computer shooting mode and that she filled in 15 hits and 15 misses and no randomly shot arrows. Moreover, she chose the user mode with respect to the game rule and filled in 15 points added for a hit and 15 point less for a miss. After shooting she did not remove any arrows. In event 2, she changed the game rule to 30 points added in case of a hit. In event 4, she found that 16 added for a hit and 15 less for a miss resulted in the correct solution. In event 5, she applied this knowledge to find another correct solution.

Table 1 displays the complete overview of the student's working as reflected in the log file in Figure 6. In this table, we also provide the total points in order to help the reader follow the student's activity.

Table 1
Student's working as reflected in the log file

Problem	Event	# hits	# misses	Game rule		Total points
				Points per hit	Points per miss	
15 hits-15 misses-15 points	1	15	15	+15	–15	0
	2	15	15	+30	–15	225
	3	15	15	+15	–16	–15
	4	**15***	**15**	**+16**	**–15**	**15**
	5	**15**	**15**	**+17**	**–16**	**15**
16 hits-16 misses-16 points	1	**16**	**16**	**+17**	**–16**	**16**
	2	**16**	**16**	**+18**	**–17**	**16**
100 hits-100 misses-100 points	1	**50**	**50**	**+2**	**–1**	**50**
	2	10	10	+4	–2	20
	3	**10**	**10**	**+4**	**–3**	**10**

* The bold printed figures belong to a correct solution.

session: 1 date: 2008/11/11 08:57:10 duration: 00:05:57 total events: 5

event: 1
who shoots: computer hits: 15 misses: 15 at-random: 0 removed: 0 *Cancel-out*
game rule: student hit: 15 added miss: 15 less

event: 2
who shoots: computer hits: 15 misses: 15 at-random: 0 removed: 0
game rule: student hit: 30 added miss: 15 less

event: 3
who shoots: computer hits: 15 misses: 15 at-random: 0 removed: 0 *Trial-and-error*
game rule: student hit: 15 added miss: 16 less

event: 4
who shoots: computer hits: 15 misses: 15 at-random: 0 removed: 0
game rule: student hit: 16 added miss: 15 less

event: 5
who shoots: computer hits: 15 misses: 15 at-random: 0 removed: 0 *Applying a general rule*
game rule: student hit: 17 added miss: 16 less

session: 2 date: 2008/11/17 08:50:24 duration: 00:04:20 total events: 3

event: 1
who shoots: computer hits: 16 misses: 16 at-random: 0 removed: 0 *Applying a general rule*
game rule: student hit: 17 added miss: 16 less

event: 2
who shoots: computer hits: 16 misses: 16 at-random: 0 removed: 0 *Applying a general rule*
game rule: student hit: 18 added miss: 17 less

event: 3
who shoots: student hits: 2 misses: 0 at-random: 1 (misses: 1) removed: 0
game rule: student hit: 99 added miss: 98 less

session: 3 date: 2008/11/17 08:54:48 duration: 00:08:53 total events: 4

event: 1
who shoots: computer hits: 16 misses: 16 at-random: 0 removed: 0
game rule: student hit: 17 added miss: 16 less

event: 2
who shoots: computer hits: 50 misses: 50 at-random: 0 removed: 0 *Analogous problem*
game rule: student hit: 2 added miss: 1 less

event: 3
who shoots: computer hits: 10 misses: 10 at-random: 0 removed: 0 *Analogous problem*
game rule: student hit: 4 added miss: 2 less

event: 4
who shoots: computer hits: 10 misses: 10 at-random: 0 removed: 0 *Analogous problem* *Trial-and-error*
game rule: student hit: 4 added miss: 3 less

Figure 6. Student's log file and applied strategies (*in italic*)

By looking at students' activities, we could identify the strategies they applied when solving the "*n* hits, *n* misses, and *n* points" task. Table 2 gives an overview of the strategies.

Table 2
Overview of strategies

Strategy	Description	Example
Altering/Ignoring information	The student arrived at the targeted result, but only part of the information was used.	The student shot 15 hits and 0 misses and assigned 1 point to each hit.
Using extreme values (Reducing complexity)	The student assigned the value 0 to one of variables so that the other variable could get the maximum value.	The student used the game rule +1 (1 point added for hit) –0 (0 points less for a miss).
Erroneously derived rule	Based on a correct answer to a problem the student applied an erroneously derived rule to provide more answers to this problem.	Based on the correct game rule +2 –1 the student applied the game rule +4 –2 (i.e., ratio of *points per hit*: *points per miss* is 2).
Repeating	The student repeated a (correct) answer to a problem to provide more answers to this problem.	
Trial-and-error	After one or more trials the student came up with a (correct) answer.	The student applied several game rules until she came up with the targeted result.
Splitting the problem*	The student answered a problem in two steps and added the partial scores to calculate (mentally) the total score.	The student shot first 100 hits and 0 misses, then 0 hits and 100 misses, and added the two partial results.
Systematic trialing	The student adjusted systematically the numbers involved in the solution until a (correct) answer was found.	The student applied consecutively the game rules: +6 –3, +6 –4, +6 –5.
Reversing solution	The student reversed a correct answer to a problem to provide more answers to this problem.	The student first applied the game rule +2 –1 and then the rule –1+2.
Transposing variables*	The student exchanged the values of arrows and points.	The student shot 2 hits and 1 miss and used the game rule +100 –100.

Solving an analogous* problem	The student substituted the numbers of a problem with smaller numbers.	Instead of shooting 100 hits and 100 misses the student shot 10 hits and 10 misses.
Cancelling out	The partial (negative) score of the misses cancels out the partial (positive) score of the hits. The total score becomes 0.	The student shot 15 hits and 15 misses and applied the game rule +1 –1.
Applying a general rule	The student applied a general solution.	The student applied the general rule where the sum of points per hit and points per miss is 1, e.g.: +100 –99.

*These strategies are induced by the built-in constraint of the maximum number of arrows that is 150.

By means of this overview of strategies, we could reveal the development in the student's solution process as it was reflected in the log file included in Figure 6. First, the student applied a trial-and-error strategy for finding a local solution. Then, she applied a general rule for solving a problem with an identical structure. Surprisingly, later on she reverted to the trial-and-error strategy.

4.3 *Advantages and concerns related to using screen videos and log files*

As shown in the examples in this section, on the one hand, interpreting students' mathematical activity is more complicated and time consuming than marking a response as correct or incorrect. But, on the other hand, the log files and the screen videos provide a detailed access to the solution process and reveal much more than a single summary score does. Such process information can be used for diagnostic purposes and also to assess students' level of understanding. For example, the log files can reveal that the problem solving strategies that students apply differ in the degree of sophistication.

In contrast with oral assessment, in ICT-based assessment this information can be provided without the presence of the interviewer and can even take place from a distance. Therefore, we coin the term "tele-assessment" for this way of assessing students. Characteristic of this assessment is that it can facilitate both small-scale and large-scale assessment. Therefore, the computer makes it possible for large-scale assessments to be focused not only on answers but also strategies.

Furthermore, technology provides students with increasing opportunities to receive feedback almost immediately thereby enabling them to revise and improve their understanding and reasoning (Kolovou and Van den Heuvel-Panhuizen, 2010). In addition, feedback can elicit students' reflection and self-assessment. The latter makes clear that revealing students' thinking and solution processes touches not only the perspective of the assessor but also that of the student: ICT-based assessment gives the students access to their own thinking processes.

Apart from these unique advantages of ICT-based assessment, there are also two issues which need to be taken into consideration when applying this assessment format.

First, our examples have revealed that the computer can record a vast amount of data; however, choices need to be made about which features of a student response have to be isolated for scoring purposes. In addition, these features have to be connected to inferences about performance (see also Bennett and Persky, 2002). For example, in the case of the log files, one has to think about what aspects of the students' mathematical activity need to be registered. This, in fact, defines the validity of the assessment, in other words, to which degree the instrument fulfills the aims of the assessment.

Second, only a range of the students' thinking is shown through the interactions of the students with the computer (Burkhardt and Pead, 2003). For example, in the case of working with screen video data, crucial steps in the student's solution process may not be captured. A combination of using screen videos and think aloud protocols may be more suitable to assess the students' strategies.

5 Final Remark

In this chapter, the focus was on how ICT can be used to improve assessment. We generally see three ways in which ICT can contribute towards better assessment. In our view, ICT can bring in tasks that have high mathematical demand, can make tasks more accessible for students, and can reveal students' thinking and solution processes. Each of these approaches has great potential to add value to learning and teaching of mathematics. We hope this is illustrated by the examples that we have presented in this chapter and wish that they may be an inspiration for those involved in thinking about new approaches in assessing students' development in mathematics.

References

Assessment Reform Group (2002). *Assessment for Learning: Research-based principles to guide classroom practice*. Retrieved February 6, 2011, from www.assessment-reform-group.org/CIE3.PDF

Barmby, P., Harries, T., Higgins, S. & Suggate, J. (2009). The array representation and primary children's understanding and reasoning in multiplication. *Educational Studies in Mathematics, 70*(3), 217-241.

Bennett, R.E. (1998). Reinventing assessment: Speculations on the future of large-scale educational testing. Princeton, NJ: Policy Information Center, Educational Testing Service. Retrieved February 15, 2011, from http://www.ets.org/Media/Research/pdf/PICREINVENT.pdf

Bennett, R.E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment 6* (9). Retrieved January 26, 2011 from http://www.jtla.org.

Bennett, R. E., & Persky, H. (2002). Problem solving in technology-rich environments. In C. Richardson (Ed.), *Assessing gifted and talented children* (pp. 19-33). London: Qualifications and Curriculum Authority.

Black, P., & Wiliam, D. (1998a). Inside the black box. Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Black, P., & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Bottge, B.A., Rueda, E., Kwon, J. M., Grant, T., & LaRoque, P. (2007). Assessing and tracking students' problem solving performances in anchored learning environments. *Educational Technology Research and Development, 57*(4), 529-552.

Bottino, R.M., & Chiappini, G. (1998). User action and social interaction mediated by direct manipulation interfaces. *Education and Information Technologies, 3*, 203-216.

Burkhardt, H., & Pead, D. (2003). Computer-based assessment: A platform for better tests? In C. Richardson (Ed.), *Whither assessment?* (pp. 133-148). London: Qualifications and Curriculum Authority.

Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology, 33*(5), 593-602.

Clements, D. (2002). Computers in early childhood mathematics. *Contemporary Issues in Early Childhood, 3*(2), 160-181.

Clements, D.H., & McMillen, S. (1996). Rethinking concrete manipulatives. *Teaching Children Mathematics, 2*(5), 270-279.

De Lange, J. (1992a). Critical factors for real changes in mathematics learning. In G. C. Leder (Ed.), *Assessment and learning of mathematics* (pp. 305-329). Hawthorn, Victoria: Australian Council for Educational Research.

De Lange, J., (1992b). Assessment: No change without problems. In W. M. Stephens & J. F. Izard (Eds.), *Reshaping assessment practices: Assessment in mathematical sciences under challenge* (pp. 46-76). Victoria: Australian Council for Educational Research.

Elbaum, B. (2007). Effects of an oral testing accommodation on the mathematics performance of secondary students with and without learning disabilities. *The Journal of Special Education, 40*(4), 218-229.

Helwig, R., Rozek-Tedesco, M., & Tindal, G. (2002). An oral versus a standard administration of a large-scale mathematics test. *The Journal of Special Education, 36*(1), 39-47.

Janssen, J., Scheltens, F., & Kraemer, J. (2005). *Leerling-en onderwijsvolgsysteem. Rekenen-wiskunde groep 4. Handleiding* [Student and education monitoring system. Mathematics Grade 2. Teachers guide]. Arnhem, the Netherlands: CITO.

Kolovou, A., & Van den Heuvel-Panhuizen, M. (2010). Online game-inherent feedback as a way to support early algebraic reasoning. *International Journal of Continuing Engineering Education and Life Long Learning, 20*(2), 224-238.

Nathan, M. J. (1998). Knowledge and situational feedback in a learning environment for algebra story problem solving. *Interactive Learning Environments, 5*(1), 135-159.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what students know. The science and design of educational assessment*. Washington, DC: National Academy Press.

Peltenburg, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (2010). ICT-based dynamic assessment to reveal special education student' potential in mathematics. *Research Papers in Education, 25*(3), 319-334.

Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders*. Needham, MA: Allyn and Bacon.

Ripley, M. (2003). The e-assessment tail. New opportunities? New challenges? The uses of ICT in assessment. Summary of presentation, 19 November 2003, Watershed Media Centre, Bristol. Retrieved January 26, 2011, from http://www2.futurelab.org.uk/resources/documents/event_presentations/beyond_exam/martin_ripley_paper.pdf

Roschelle, J. M., Pea, R. D., Hoadley, C. M., Gordin, D. N., & Means, B. M. (2000). Changing how and what children learn in school with computer-based technologies. *The Future of Children, 10*(2), 76-101.

Sarama, J., & Clements, D. H. (2006). Mathematics, young students, and computers: Software, teaching strategies and professional development. *The Mathematics Educator, 9*(2), 112-134.

Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne & M. Scriven (Eds.), *Perspectives on curriculum evaluation* (AERA Monograph Series – Curriculum Evaluation). Chicago: Rand McNally and Co.

Shepard, L. A., (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Van den Heuvel-Panhuizen, M., & Becker, J. (2003). Towards a didactic model for assessment design in mathematics education. In A.J. Bishop, M.A. Clements, C. Keitel, J. Kilpatrick, & F.K.S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 689-716). Dordrecht: Kluwer Academic Publishers.

Van den Heuvel-Panhuizen, M. (2007). Changes in purpose, content and format: Trends in assessment in mathematics education. In B. Choksi & C. Natarajan (Eds.), *The epiSTEME Reviews, Volume 2, Research trends in science, technology and mathematics Education* (pp. 241-266). New Delhi, India: Macmillan.

Vygotsky, L.S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Chapter 9

The Assessment *for*, *of* and *as* Learning in Mathematics: The Application of SLOA

Magdalena Mo Ching MOK

Globalisation and knowledge economy demand a renewed vision on pedagogy. This chapter presents a framework of learning and assessment integration for pedagogical re-vision in the context of mathematics education. This framework, termed the Self-directed Learning Oriented Assessment (SLOA), is grounded on cognitive learning theory and underpinned by the belief that all assessment activities should contribute to learning. In SLOA, the learning-assessment integration is considered from three perspectives. First, Assessment *of* Learning informs the learner about how much has been learned, and identifies the gap between intended learning goal and current achievement. Second, in Assessment *for* Learning, assessment is a vehicle for informing the learner how to enhance future learning. That is, feedback from assessment is used to feedforward. Third, Assessment *as* Learning means the learner internalises assessment as part of learning and becomes a self-directed learner. The SLOA framework encourages student participation, and highlights the role of self-evaluation and self-assessment. By returning ownership of learning to the student, the framework makes assessment a transparent process and consequently accountable at all levels of education.

1 Introduction

In response to rapid changes in education, resultant from globalisation, technological advancement, and the changing nature of work, there has been increasing emphasis in the 21st century by key education systems in the preparation of students who are *trainable* rather than *trained*, to become *self-directed* rather than *teacher-directed* learners. There is strong consensus among governments that building capacity for self-directed learning and sustaining motivation to learn are key education pursuits. For instance, the vision of education for Singapore is to develop self-directed learners who take responsibility for their own learning, who question, reflect and persevere in their pursuit of learning (Ministry of Education, 2010). Resources have been invested by developed countries worldwide to promote competencies for learning throughout life. The centrality of the capacity for self-directed learning is underscored as essential for the implementation of the 1996 "lifelong learning for all" vision of the Organisation for Economic Co-operation and Development (OECD) Education Ministers (OECD, 2004). In parallel, there has been a proliferation of research in recent decades on self-directed learning with significant impact on learning and instruction (Dignath and Büttner, 2008; Winne, 2005).

Investigation into self-directed learning is particularly necessary in Asian countries like Singapore and Hong Kong for several reasons. First, there is strong societal and familial expectation in these countries for children to succeed academically. Outcomes are normally attributed to effort rather than ability (Phillipson, 2006). Second, classrooms of these countries are often portrayed as highly competitive, teacher-centred, large-group teaching, and places where learning is characterised by transmission of declarative knowledge rather than knowledge construction (Kan, 2010). This environment is unfavourable for the development of self-directed learning competencies. Third, research shows that not all research findings in the West are directly transferrable to learners of the other cultures. For example, Lee, Yin, and Zhang (2009) found that, in Hong Kong teacher rather than student involvement and support in classrooms was the strongest predictor of self-regulated learning. Fourth, there is ample research evidence that students who are self-directed in their learning also achieve better (Schunk, 1998).

Nevertheless, although positive effect of self-directed learning on academic outcomes has been reported by researchers and findings tend to be replicated across North American and Western Europe, few studies were undertaken in Asian countries (Law, Chan, and Sachs, 2008; Lee, Yin, and Zhang, 2009). Further, whereas conceptually self-directed learning has been embraced at the policy level, a review of recent literature suggests that implementation at the practical level is limited. There is little evidence of strategies that support teachers putting self-directed learning theories into practice (Dignath and Büttner, 2008; Stoeger and Ziegler, 2008).

In this chapter, I draw on classroom research undertaken in China, Hong Kong, and Macau concerning the successful implementation of a model for the development of self-directed learning in primary and secondary students. The model is called Self-directed Learning Oriented Assessment (SLOA). Discussion in this chapter focuses attention on strategies of applying SLOA in the context of mathematics learning in primary and secondary classrooms.

In this chapter, I discuss how assessment can be used to foster self-directed learning. The use of assessment as a means to promote the capacity to take charge of one's own learning is called Assessment *as* Learning by Earl (2003). Unlike in the past when assessment was used solely for selection purposes, assessment takes on several functions in the 21st century. Another function of assessment is the generation of feedback to promote metacognition and enhance achievement, or what is known as Assessment *for* Learning. Extant research shows that feedback is a powerful factor influencing learning and achievement (Black and Wiliam, 1998, 2009; Hattie and Timperley, 2007). Feedback provides students with information about their learning, and the gap between current and intended performance. Students can then readjust their learning strategies or refine learning goals. In order to ascertain where the students have reached in comparison to intended achievement, Assessment *of* Learning is necessary. Assessment *as* Learning, *for* Learning, and *of* Learning are the three components of SLOA.

SLOA is a coherent framework of assessment, deliberately designed to capitalise on the integrative impact of metacognition, feedback, motivation, contextual factors, and self-regulation on learning in the construction of assessment activities in order to cultivate self-directed learning capacities in students. The SLOA framework has been informed by previous research in the area of learning psychology, particularly research in feedback, metacognition, assessment, and self-directed learning (Black and Wiliam, 1998, 2009; Boekaerts and Cascallar, 2006; Carless, 2007; Flavell, 1987; Hattie and Timperley, 2007; Hogan and Gopinathan, 2008; Schunk, 2008; Schunk and Zimmerman, 2007; Weiner, 1986; Zimmerman, 2008). The essence of SLOA is the notion that assessment should serve and advance learning, and that the learner should be self-directed. Three principles are fundamental to SLOA: (a) assessment should comprise learning tasks; (b) assessment should be designed to engage students in learning; (c) feedback generated from assessment should feedforward to inform subsequent learning (Carless, 2007). In the SLOA framework, evidence of learning is conscientiously gathered in order to develop meaningful connections among the student's learning progress, his/her level of achievement, misconceptions, learning difficulties, short- and long-term learning goals, curriculum, and assessment such that the next phase of teaching and learning can be more effective.

2 Implementation of SLOA in Mathematics

The SLOA framework can be applied to all curriculum subjects but its effective implementation is best to be contextualised. Mathematics is a good subject for contextualising SLOA because mathematics involves deep and systematic thinking, and research has shown that mathematics learning benefits from developing students' metacognition.

2.1 *Building a SLOA school culture: shape mindsets and change practice*

Mathematics learning does not happen in a vacuum. For effective mathematics learning, supportive school and classroom learning environment is crucial. Translated to SLOA terminology, this means to have a whole school approach, build a culture of SLOA in schools through shaping mindsets and changing practice. Zimmerman (2008) put forward a cyclical model of self-directed learning, comprising three phases, namely (a) the forethought phase during which the learner analyses the task, sets learning goals, and plans the study; (b) the performance phase during which the learner exercises self instruction, focuses attention, enacts learning strategies, exercises metacognitive monitoring, and self records learning outcomes, and (c) the self-reflection phase during which the learner self-evaluates, makes causal attribution, and self-regulates to enhance subsequent performance. A self-directed learning school engages itself in these three phases of self-directed learning conscientiously. In other words, school leaders should set the learning goals in mathematics for the whole school so that there is a coherent mathematics curriculum across year levels, develop pedagogy for effective mathematics teaching and learning, establish assessment strategies at the systemic level, undertake systematic self-review to enhance group metacognition, instigate mechanisms for school self-regulation to enhance overall achievement.

2.2 *Professional development: equipping mathematics educators for the effective implementation of SLOA*

Recent research (e.g., Boekaerts and Corno, 2005) found a shift in professional development from a strategy training approach to an integration of training on both domain-specific knowledge and knowledge about learning. Professional development of SLOA is more effective if it is embedded in a school subject such as mathematics.

Further, innovation is usually associated with change-induced stress. Increased workload particularly at the start of the intervention, apprehension about the lack of technical knowledge for the implementation, fear of exposure to criticisms by parents and peers, and lack of confidence in the efficiency and effectiveness of the new method are the common stressors. Support from school leaders will reduce the stress (Kayler, 2009). In particular, a safe work environment where the teachers can feel secure to experiment with the new approach (Day, 2008), alignment of teachers' goals with the mission of the school (Chan, Lau, Nie, Lim, and Hogan, 2008), and provision of hard evidence on students' increase in achievement (Day, 2008) are strategies to enhance teachers' willingness to implement innovations (Ballet and Kelchtermans, 2009).

For the implementation of SLOA in mathematics it is critical that teachers are able to set assessment tasks that discriminate the performance of students. For this purpose, mathematics teachers find teacher development in item setting particularly useful. Two recent developments of item setting are worth mentioning here and the first concerns the use of two-tier items to detect areas of weak understanding. Two-tier items are testlets with two related items assessing the same concept bundled together. The first item is usually designed to assess declarative knowledge and the second to explore reasons behind students' responses to the first item. Two-tier items have the benefits of revealing not only whether or not the students have mastered a concept, but also where misconceptions have occurred. Since it is very difficult to get both items correct by guessing, two-tier items can overcome some of the drawbacks of multiple choice items due to guessing. An example of a two-tier item is presented in Figure 1.

1. Which of the two fractions 3/5 and 2/7 is larger?
   A. 4/5 is larger than 2/7
   B. 2/7 is larger than 4/5
   C. 2/7 and 3/5 are equal in size
2. Please give the reason for your choice in Item 1.
   A. 4 is greater than 2, so 4/5 is larger than 2/7
   B. 4 is greater than 2, so 2/7 is larger than 4/5
   C. 7 is greater than 5, so 4/5 is larger than 2/7
   D. 7 is greater than 5, so 2/7 is larger than 4/5
   E. 4 is larger than 2 by 2, and 7 is larger than 5 by 2, so they are equal.

Figure 1. Example of a two-tier item in fractions

A second new development in item setting relates to cognitive diagnostic assessment (de la Torre, 2009). Cognitive diagnostic models are psychometric models for the diagnosis of students' strengths and weaknesses in their learning. In this branch of study, assessors have to first of all establish the knowledge base of the topic to be tested, delineate the attributes underpinning the topic, and then set items to capture students' degree of mastery of these attributes. The oft quoted example of cognitive diagnostic assessment is work on subtraction of fractions where five underpinning attributes are identified (Tatsuoka, 2009). The five attributes are illustrated by the example in Figure 2.

		Attribute				
Item		1	2	3	4	5
1	$\frac{3}{5}-\frac{1}{5}$	✓	×	×	×	×
2	$\frac{2}{3}-\frac{1}{6}$	✓	✓	×	×	×
3	$2\frac{1}{5}-3\frac{6}{5}$	✓	✓	✓	✓	✓
4	$5-\frac{3}{4}$	✓	×	✓	✓	✓
… etc…						

Attributes:

1. Performing basic fraction subtraction operation
2. Simplifying/reducing
3. Separating whole number from fraction
4. Borrowing one from whole number to fraction, and
5. Converting whole number to faction.

Figure 2. Attributes of Fraction-subtraction

Item 1 in Figure 2 requires attributes 1, and Item 2 requires attributes 1 and 2. If the items are set thoughtfully, then by simply comparing performance across items, the teacher can identify areas of strength and weaknesses. For more sophisticated problems, students' assessment data can be analysed using the DINA (deterministic, inputs, noisy "and" gate) model (de la Torre, 2009) to find out exactly where the difficulties lie for each student taking the assessment.

3 Assessment *for* Learning in Mathematics

3.1 *Pedagogical approaches for the implementation of Assessment for Learning in mathematics*

Assessment *for* Learning is central to SLOA. It refers to the generation of feedback from assessment in order to inform and advance learning. Feedback can be generated from monitoring processes that inform learners about the standards of the task, the match or gap between the desired goal and actual achievement, and the effectiveness of learning strategies.

The types of feedback, the way feedback is given, and the way feedback is received all affect the power of feedback on learning (Hattie and Timperley, 2007). Feedback that is directed at the self or personal level (e.g., in the form of personal praise or criticism) tends to be debilitating as the learner subsequently would minimize the risk to the self by avoiding challenging tasks or failure (Black and Wiliam, 1998; Hattie and Timperley, 2007). Feedback is formative or facilitates learning when it is directed towards the quality of the task or learning process, especially when it helps to identify misconceptions and then leads to clarification of concepts, or development of more effective learning strategies (Hattie and Timperley, 2007). Feedback is most effective when it helps to identify the gap between the learner's performance and his/her learning goal, and then provides advice on how to improve from there (Black and Wiliam, 1998; Hattie and Timperley, 2007). When the gap is clearly identified, feedback can increase effort and motivation to close the gap, and/or to search for clues and engage in task processes that increase understanding.

Effective self-directed learners make use of both internal feedback generated from self-monitoring and self-reflection, and external feedback from teachers or peers to optimise cognitive awareness and academic outcomes (Labuhn, Zimmerman and Hasselhorn, 2010; Paris and Paris, 2001). External feedback from teachers can act as a basic standard and reference point and is particularly useful for students' metacognitive processes including goal-setting, self-monitoring, and self-evaluation (Paris and Winograd, 1990).

Hattie and Timperley (2007, p. 86) recommended three questions that students should form the habit of asking themselves to generate self-feedback:

(1) "Where am I going", or what is the desired outcome (long-, intermediate-, short-term) of my learning endeavour? What is the anticipated outcome if I approach the problem this way? How is this new learning related to my previous learning?

(2) "How am I going", or what does the assessment evidence tell me about the effectiveness of my learning strategies and is there a gap between my desired goal and my current progress? If there is a gap, what are the possible causes? And,

(3) "Where to next", or what should be my next steps? Do I have to keep going this way or should I modify my learning strategies? Should I change my goal (set higher/lower goal; change direction)? Should I seek help and if so from where should I get help?

Assessment *for* Learning takes time to develop and for the newcomers, the process can be daunting: students expose themselves to analyses by themselves, peers and the teacher on their deficiencies and misconceptions. Care must be taken in building a classroom environment in which students feel safe enough to experiment with the new experience. The purpose of feedback must be explained clearly to the students at the outset. Further, rules, protocols, and etiquettes have to be established and agreed upon by all those involved. For instance, teachers may want to reach consensus with students on the following:

- Feedback has to aim at improvement of learning;
- Feedback has to focus on issues rather than on person;
- Give reasons and explain why you disagree or agree with the other student;

- Point out explicitly any possible mistake or misconception the other person may have;
- Acknowledge other students' contributions before putting in a different view;
- Compare and contrast different approaches to solving an item;
- Everyone has to contribute to the group learning in some way; and
- Raise your hand if you want to express an idea, and speak in turn.

Teachers may want to construct some "standard" statements at the beginning as scaffolding for those weaker in language and expression, and for younger students. Our experience shows that even children as young as primary 2 can learn how to provide thoughtful feedback. Further, students enjoy class discussion and peer feedback if they can see the benefits of these activities. As one student wrote in her reflective journal, "I like class discussion because I can ask for help if I don't understand, and I can help others too".

Feedback in Assessment *for* Learning helps both learning and teaching. In order to generate feedback from assessment to inform teaching, teachers have to design informative assignment tasks. Designing and using two-tier items is one approach, but effective use of multiple choice items is another effective method. The key lies in making predictions of students' response before administering the assessment items. Teachers should deliberate upon the common mistakes and difficulties encountered by capable students, weak students and the averaged student. For instance, an item on estimation is presented in Figure 3. Common mistakes in estimation are: (a) rounding up the two numbers before addition. This common mistake is incorporated in Option A; (b) Rounding up or rounding down component numbers before addition. This common mistake is incorporated in Option B; (c) No idea about what is estimation at all. This common mistake is incorporated in Option D. The remaining option, Option C, represents the correct procedures for estimation.

Estimate the value of 342.7 + 58.51

A. 402
B. 401
C. 400
D. 401.21

Figure 3. A Multiple Choice item with options to reveal common misconceptions

Misconception is only one source of error for students. Nevertheless, difficulties faced by students in problem solving may not be confined to misconceptions only. Students may be hindered by language issues, for instance. Teachers should set items and include problem solving strategies to help delineate where the problem lies if a student gets an item wrong. For instance, the item in Figure 4 is set to test students' understanding of "equal shares" in the concept of fractions but students may get the item wrong for a number of reasons: they do not know the word fraction, so they just counted the number of shaded parts (Option A); cognitive load may be too high for some students and they cannot mentally hold all the different parts together, so they just focus on one shaded part within one of the smaller squares (Option B); they may not recognise all the parts have to be equal in size for a fraction, so they count the number of shaded pieces, use it as the numerator, count the total number of pieces and use it as the denominator (Options C or E; depending on whether they simplify or not). Only those who can visualise and move one of the shaded parts to be adjacent to the other shaded part get the correct option (Option D). Teachers can generate information about the difficulties faced by students using carefully crafted options in multiple choice items. In order to find out whether students understand the English words "fraction", "shaded", and "parts", teachers may administer a simple item, such as the item in Figure 5, first of all.

What is the fraction representing the shaded parts in figure A?

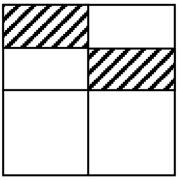


Figure A

A. 2 B. $\frac{1}{2}$ C. $\frac{1}{3}$ D. $\frac{1}{4}$ E. $\frac{2}{6}$

Figure 4. Representation of Shaded Parts in a figure

What is the fraction representing the shaded part in figure B?

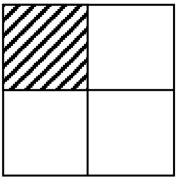


Figure B

A. 1 B. 4 C. $\frac{1}{3}$ D. $\frac{1}{4}$ E. $\frac{3}{4}$

Figure 5. Representation of Shaded Parts in a figure

3.2 *Tools for Assessment for Learning*

The essence of Assessment *for* Learning is that assessment can be used to inform and support further learning. Assessment can be formative if teachers know where to exercise scaffolding. The Rasch model (Bond and Fox, 2007) is a tool for the identification of the Zone of Proximal Development (ZPD) (Vygotsky, 1978 in Rieber and Carton, 1987) for each student. Three "maps" outputs from the Rasch analysis are particularly useful to teachers, namely, (a) Item-Person Map; (b) Kid-Map; and (c) Unexpected Persons Map.

The Item-Person Map is a visual display of students taking the test, and the items in the test along the same measurement scale. An example is presented in Figure 6. In this Map, student abilities are arranged from the lowest (at the bottom of the map) to the highest (at the top of the

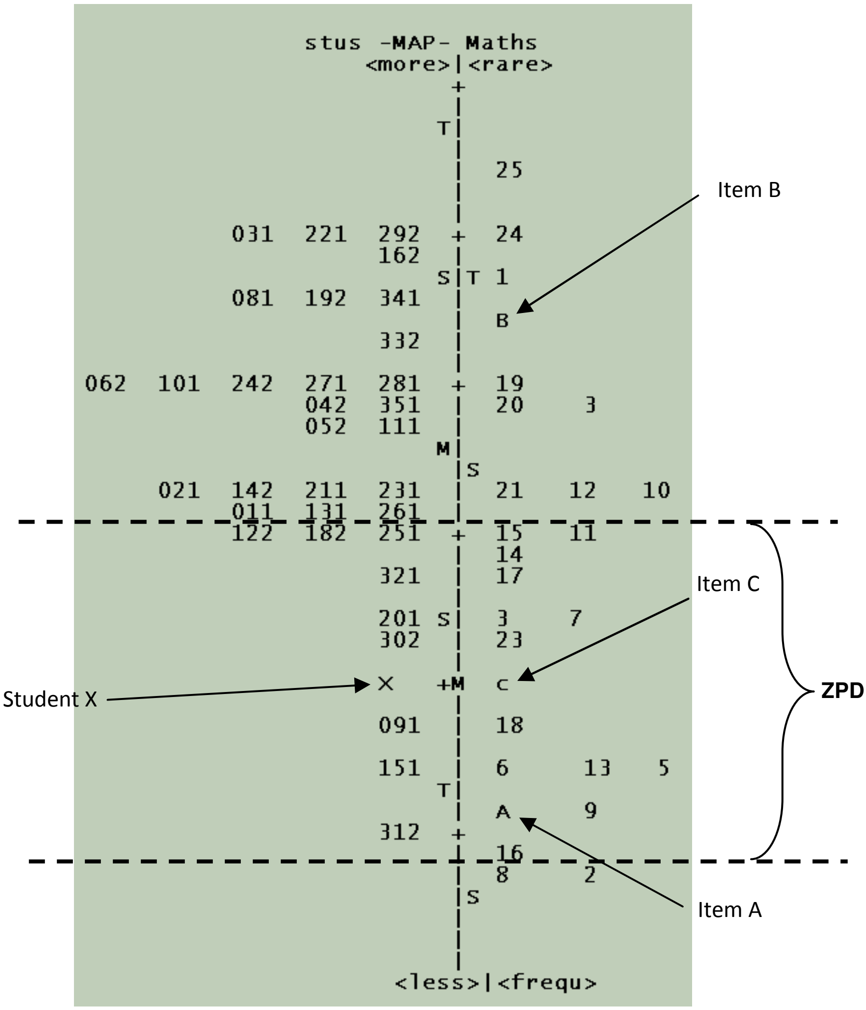


Figure 6. Item-Person Map

map), and presented on the left of the measurement scale. Items are arranged from the least difficult (at the bottom of the map) to the most difficult (at the top of the map). If the ability of a student (e.g. student X) is more than the level of difficulty required to solve an item (e.g. item A), then chances are high for the student to get the item correct. On the map, this is indicated by the student being above the item on the display.

If on the other hand the ability of the student (e.g. student X) is lower than the level of difficulty required to solve an item (e.g. item B), then chances are high for the student to fail the item. On the map, this is indicated by the student being below the item on the displayed. If the ability of the student is equal to the difficulty level of the item (e.g. Item C), then the student has 50% chance to get the item correct. By drawing horizontal lines for each student, the teacher can find out the region where the student has 50% chance of getting the item correct. This is the Zone of Proximal Development of the student. Around this region, the student has 50-50 chance of getting the items correct but with scaffolding from the teacher, the student can improve his/her chance of getting right answers to beyond guessing.

The Kid-Map provides information on expected and unexpected response patterns for each student taking an assessment. The Kid-Map has four quadrants and the students' ability level is represented by "XXX" in the map (Figure 7). On the left are items attempted correctly by the student and on the right are those items that the student got them wrong. Each side is in turn divided into the top and bottom quadrants. The top quadrants (Harder Achieved, and Harder Not Achieved quadrants) represent those items with difficulty levels beyond the ability level of the student, and items at the bottom quadrants (Easier Achieved, and Easier Not Achieved quadrants) are below the student's ability. For instance, student got item 30 correct (Figure 7). This is in fact a difficult item, more difficult than the student's ability level. The teacher may want to find out why this has happened. Is there some untapped potential of the student? Has the student answered the items honestly? Is there any component of guessing or luck? More importantly, items 5, 47, 50, 45, 46, 23, 4, 27, and 49 are items below the ability level of the student and the student was expected to answer them correctly but the student got them wrong. What are the concepts, skills, and attributes underpinning these items that the student has not mastered? These items represent areas for remediation. The items in the Easier Achieved quadrant (e.g. item 40) represent areas where the student has mastered and the teacher can use the areas as foundation to build up the student's relevant mathematics ability. The region one standard deviation above and below the student's ability level is indicated by "……" above and below the

"XXX" in the map, and contains items 41, 29, 13, 8, 22, 37, and 11. This could be taken as the Zone of Proximal Development of the student. This is where scaffolding can be applied.

The Unexpected-Person Map gives teachers an in-depth understanding of the errors made by individual students. An example for student number 27 is presented in Figure 8. Person ability (from 10 to 100 in the Figure) and item difficulty are presented as the horizontal and vertical axis respectively. Item Q23 (with entry number 33) is the most difficult, and item Q2 (with entry number 2) is the least difficult, item in the list. The student has ability measure 74.86 estimated using the Rasch model. For persons at this ability level, they are expected to fail Q23 (represent by "0" in the map), but to get items Q11, Q10, Q4 and Q2 correct (represented by the column of "1"s in the map).

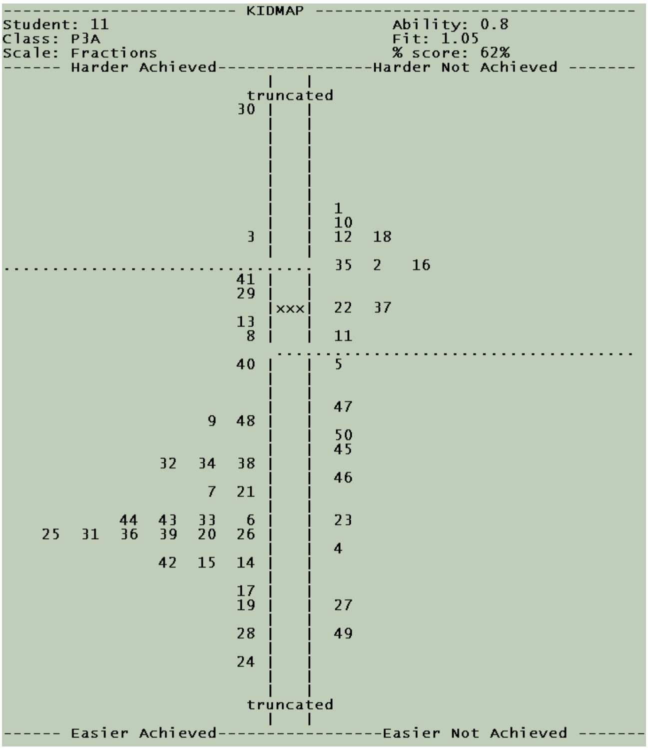


Figure 7. An example of a KIDMAP

However, it is observed that for this student, s/he has answered item Q23 (the most difficult item) correctly. Only persons with ability level around 92 are expected to get item Q23 right. On the other hand, this student has failed items Q11, Q10, Q4, and Q2 and these results are highly unexpected because only those students with ability level around 42 are expected to fail Q11; those with ability level around 38 are expected to fail Q10; those with ability level around 32 are expected to fail Q4; and those with ability 20 are expected to fail item Q20. The teacher is provided with detailed information about this student to follow up with.

```
KEY: .1.=OBSERVED, 1=EXPECTED, (1)=OBSERVED, BUT VERY UNEXPECTED.

NUMBER - NAME ------------------ MEASURE - INFIT (MNSQ) OUTFIT - S.E.
     27  27                        74.86     1.6    A     4.1      5.85

10    20    30    40    50    60    70    80    90   100
|-----+-----+-----+-----+-----+-----+-----+-----+-----|  NUM  item
                                         0        .1.       33  Q23
                  (0)                    1                  11  Q11
                (0)                      1                  10  Q10
            (0)                          1                   4  Q4
     (0)                                 1                   2  Q2
|-----+-----+-----+-----+-----+-----+-----+-----+-----|  NUM  item
10    20    30    40    50    60    70    80    90   100
```

Very unexpected response

Expected pattern of responses for students with ability Measure 74.86

Figure 8. Unexpected- Person Map

In summary, the Rasch model can be used to provide the teacher with both a global picture about the whole class, and detailed information about individual students for remedial actions. This is the essence of Assessment *for* Learning.

3.3 *Using Student-Problem Chart for Assessment for Learning*

The Student-Problem Chart (SP-Chart) was originally devised by Takahiro Sato (1980; see also Chacko, 1998; Harnisch, 1983) as an Assessment *for* Learning tool in order to exploit the interactions between students and test items for information to support instruction. The SP-Chart can be constructed by rotating the rows and columns of the student-item response matrix such that students are arranged from top to bottom of the matrix in descending order of their total scores, and items from left to right in descending order of item total scores. This is illustrated in Figure 9 with a hypothetical example involving 7 students taking an assessment of 6 items.

	Q3	Q1	Q6	Q4	Q2	Q5	Student Total
S1	1	1	1	1	1	0	5
S4	1	1	1	1	0	0	4
S6	1	1	1	0	1	0	4
S5	1	1	0	1	0	0	3
S2	0	1	1	0	0	0	2
S3	1	0	0	0	0	1	2
S7	1	0	1	0	0	0	2
Item Total	6	5	5	3	2	1	

Key:

Student Curve

Problem Curve

Figure 9. Student-Problem Chart with Student-Curve and Problem-Curve

Since after arrangement, the items listed on the left of the response matrix are easier than those on the right, it is expected that under normal circumstances, a student (e.g. S3) with 2 correct answers will get the 2 easier items (i.e., Q3 and Q1) right, and the harder items (e.g. Q6, Q4, Q2, Q5) wrong. A vertical line can then be drawn in the matrix to separate these two sets of items for this student to indicate the expected right/wrong pattern of responses for the student. Responses to the left of the vertical line are expected to be correct answers and responses to the right are expected to be wrong. This can be done for all students. Joining up all the vertical lines gives a Student curve. Similarly a Problem curve can be formed by joining the horizontal lines that indicate the expected right/wrong responses obtained for each item.

Importantly, by inspecting the SP-Chart, the teacher can detect easily anomalies in students' response. For instance, students S2, S3, and S7 in Figure 9 both have 2 correct answers but the response pattern of S3 is aberrant. Student S3 has two correct responses, one for Q3 (the easiest item) and the other one for Q5 (the most difficult item). These irregular response patterns may be due to a number of reasons, including guessing, cheating, carelessness, inadequate mastery, lack of test-taking skills, etc.

3.4 *Avoiding common pitfalls in Assessment for Learning*

Several common pitfalls should be avoided and the first of these is the confusion between Assessment *for* Learning and Continuous Assessment. Whereas the former means the provision of feedback on student learning using information generated from assessment, the latter means frequent assessment. Frequent assessment without careful planning cannot guarantee useful information can be produced that can inform learning. Too many assessments without proper feedback takes away precious instructional time and can be stressful for students. The second pitfall is over-reliance on item-level data. Teachers have to avoid falling into the trap of operating at the item-level. Instead, they should look beyond item-level data to attributes underpinning the items. Consider when a student cannot do an item "$\frac{1}{3} - 1\frac{1}{4} =$?" Instead of giving more exercises of the same type for the student, the teacher has to find out whether the reason is due to: the student cannot distinguish between fractions and number, does not know how to find the common denominator of the two fractions, cannot carry out simple operation of fractions, or other issues.

4 Assessment *of* Learning in Mathematics

4.1 *Pedagogical approaches for the implementation of Assessment of Learning in mathematics*

Assessment o*f* Learning aims to identify the achievement of the learner at the end of key learning stages (e.g., end of a learning unit). The key questions concern how much the student has learned and whether or not the standard, usually set externally, has been reached. In the SLOA framework, Assessment *of* Learning is used not only to ascertain the present level of achievement, but also to provide an external frame of reference on that achievement. This information is used to provide an anchor for curriculum design, identification of longer term goals, and ascertain whether or not there is a gap between the targeted goals and students' achievement. In this way, Assessment *of* Learning takes up an augmentative role for Assessment *for* Learning, or summative assessment for formative use (Black, Harrison, Lee, Marshall, and Wiliam, 2003), in the SLOA framework. Providing evidence with external reference will also help students to have a better understanding of where they are, or to develop metacognition; thus Assessment *of* Learning helps to support Assessment *as* Learning in the SLOA framework.

4.2 *Tools for Assessment of Learning*

A vertical measurement scale to chart student growth across time is an important tool for Assessment o*f* Learning. Traditionally schools rely heavily on raw scores and level-specific tests and examinations to measure student achievement. Unfortunately, the use of raw scores and level-specific tests/examinations makes articulation of achievement results across year levels impossible. For instance, if a student gets 80 marks in mathematics in primary 3 and 70 marks in the same subject in primary 4, does it mean that the student has regressed in his/her learning? The substantive meaning of "80 marks" also varies from subject to subject, between teachers in the same school, and across different schools. That is, we need assessment tools that enable us to accurately gauge the amount of growth in students' learning. A mathematics vertical scale that aligns with the mathematics curriculum in Hong Kong has been developed to enable teachers to chart growth across year levels (Lau, Mok, and Yan, 2009).

A mathematics vertical scale comprises an item bank of many (over thousands) mathematical items that align with the mathematics curriculum. These mathematics items have difficulty levels spanning the curriculum years from primary 1 to senior secondary levels. The scale is then calibrated using Rasch measurement method and validated with a representative sample of students from the targeted year levels. As the items in the scale are aligned with the local curriculum, students assessed using the scale can have a clear indication of the standards reached by the student taking the test. It is beyond the scope of this chapter to discuss how a vertical scale can be developed. Interested readers may want to refer to works by Patz and Yao (2007), and Kolen and Brennan (2004).

4.3 *Avoiding common pitfalls in Assessment of Learning*

Several pitfalls should be avoided and the first of these is the assumption that Assessment *of* Learning is not necessary or less important than Assessment *for* Learning in the new era of assessment reform. On the contrary, assessment has to satisfy multiple roles and one of them is to inform parents, teachers, employers, and the government how much students have learned, and whether what students have learned have reached desired standards. This function cannot be achieved without high quality Assessment *of* Learning. A second common pitfall is the assumption that Assessment *of* Learning can only take place at the end of the school year or semester. On the contrary, Assessment *of* Learning need not be confirmed to the end of learning. It can be used at any time to give teachers an overall picture of how much is learned.

5 Assessment *as* Learning in Mathematics

5.1 *Pedagogical approaches for the implementation of Assessment as Learning in mathematics*

The term Assessment *as* Learning is coined by Lorna Earl (2003). It means the engenderment of students into self-directed learners through reflecting on evidence of learning generated from assessment activities. The desired outcome of Assessment *as* Learning is to let each student take charge of his/her own learning and learn how to learn. Metacognitive self-regulation in mathematics is an important predictor of mathematics achievement at all developmental levels (Camahalam, 2006; Desoete, 2008; Desoete, Roeyers, and De Clercq, 2003; Labuhn, Zimmerman, and Hasselhorn, 2010). Desoete (2008) found that metacognitive skillfulness combined with intelligence of primary students accounts for 53% to 76% of their mathematics performance. Desoete, Roeyers and De Clercq (2003) further found that 1/3 of children with mathematical learning disabilities also had ineffective metacognitive skills.

The process of Assessment *as* Learning involves the cultivation of capacity for goal setting, self-monitoring of learning progress, self- and peer-assessment of achievement, self-motivation in face of difficulties, and self-regulation and change in order to enhance further learning where appropriate. Teachers may consider a modified infusion approach originally designed by Swartz and Parks (1994) for the implementation of Assessment *as* Learning in mathematics. The infusion method places metacognition at the core of self-directed learning processes (Paris, 2002; Paris and Winograd, 1990). Students are taught explicitly, through direct teaching and modeling, strategies for planning, goal setting, self-monitoring, self-evaluation, and self-regulation. They are coached through systematic scaffolding (e.g. using mathematics learning logs, guided questions and teachers' feedback) to reflect on their learning and to become aware of their own thinking. Self-directed learning strategies are not taught in isolation. Rather, they are embedded in the learning activities and taught across a range of primary mathematics topics (Paris and Paris, 2001).

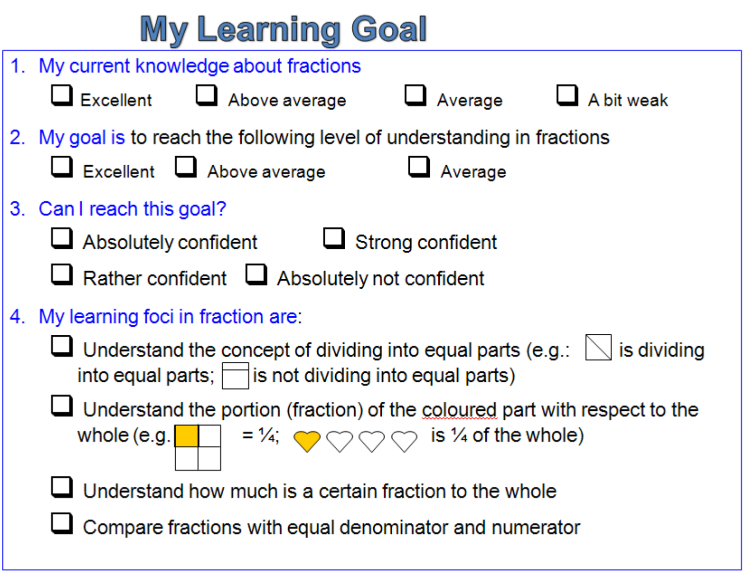
A typical activity involves the teacher presenting a problem (e.g., “Please express the fraction 10.5/500 as a percentage.”) and students are invited to present their solution (“How did you do it?” “How do you know that this is correct?”). The rest of the class is asked to present alternative solutions (“Can you solve it in any other way?”), consider merits of the solutions (“Which of these is a better approach? Why?”), reflect on future applications (“How will you tackle the problem if you meet it again?”), and to post new questions (“Can you create an item to test if another student really understands the concept?”). These and other reflection questions are revisited for other topics to build up a repertoire of self-reflection strategies. Students are actively engaged while learning mathematics in self-reflection and self-monitoring in a conscious and explicit manner both individually and as a group so that they are equipped with problem-solving skills, learning skills, and an inquiry habit of mind in context (Dewey and Bento, 2009).

The infusion approach has many reported advantages: (a) Framework: A framework is provided to the general activities that can be used by the teachers in conjunction with the mathematics curriculum where appropriate; (b) Integration: Embedding learning in context has been shown to be an effective way to learn self-directed learning (Paris and Paris, 2001; Zimmerman, 2008); (c) Flexibility: Instructional materials are flexible for teachers to fit into their teaching schedule and to align with students’ abilities; (d) Active learning: Students are engaged through activities designed for embedded instruction; (e) Spiral instruction: New self-directed learning skills are introduced gradually, with ample opportunities for practice and deepening throughout the intervention period (Spektor-Levy, Eylon, and Scherz, 2008).

5.2 *Tools for Assessment as Learning*

Critical questions for Assessment *as* Learning in the SLOA framework are: How to promote metacognition in students? How to engender and sustain their learning motivations, and how to support them in developing and internalising motivational volitional controls? In other words, how to foster self-directed learning? This is done in SLOA by teacher modeling, teacher verbal/written feedback without grades, learning log, and self-reflective journals. Vital to these strategies is to elicit reflections from students themselves. The purpose is to make subconscious processes overt such that the learner can have better control and access to them at a later stage (Black and Wiliam, 2009; Vermeer, Boekaerts, and Seegers, 2001).

A well-designed mathematics learning-log can be used to develop the habit of mind for self-directed learning. Teachers may want to make use of learning-logs to let students self-assess, and generate feedback to inform their own learning. In so doing, assessment becomes an integral part of learning. Learning logs are also good venues for parents, peers and the teacher to communicate about student learning in a non-threatened manner. Presented in Figure 10 is an example of page 1 of one such learning-log for mathematics.



My Learning Goal

1. My current knowledge about fractions
   ❑ Excellent ❑ Above average ❑ Average ❑ A bit weak
2. My goal is to reach the following level of understanding in fractions
   ❑ Excellent ❑ Above average ❑ Average
3. Can I reach this goal?
   ❑ Absolutely confident ❑ Strong confident
   ❑ Rather confident ❑ Absolutely not confident
4. My learning foci in fraction are:
   ❑ Understand the concept of dividing into equal parts (e.g.: is dividing into equal parts; is not dividing into equal parts)
   ❑ Understand the portion (fraction) of the coloured part with respect to the whole (e.g. = ¼; is ¼ of the whole)
   ❑ Understand how much is a certain fraction to the whole
   ❑ Compare fractions with equal denominator and numerator

Figure 10. Example of Mathematics Learning Goal
Courtesy Ms Doris C.H. Lau (The Hong Kong Institute of Education) and
Ms Jenny Li (Chai Wan Faith Love Primary School)

In Figure 10, students were guided to self-assess on their current knowledge, then set a learning goal in broad terms ("My goal is to reach the following level of understanding in learning fractions: Excellent, Above average, Average"), evaluate on self confidence, and set more specific goals ("My learning foci in fractions are…"). The combination of broad and specific goals helps develop students' metacognition. The specific goals are helpful for students to check against their learning outcomes and identify the gap between desired goal and actual achievement. After each assessment, students are guided using the learning-log to evaluate their own achievement. In the learning-log, they are invited to (a) assess whether they have or have not achieved their learning goal, (b) evaluate the amount of effort they put into the study, (c) make attributions for their learning outcomes, and (d) propose strategies on how to improve on their outcomes the next time.

Of particular note is the use of learning-logs in developing students' attribution for academic success or failure. Academic attribution is the explanation an individual gives to explain the reasons for their academic performance at school – successful or otherwise – after it has occurred (Weiner, 1986). Ability, effort, luck, and task difficulty are common reasons that students use in explaining academic successes and failures (Chan and Moore, 2006). Attributions have been identified as powerful determinants of student learning, achievement and self-esteem (Paris and Paris, 2001; Zimmerman and Kitsantas, 1999; Zimmerman 2000). Recent research emphasis has been placed on the use of strategies in combination with effort (e.g., Chan and Moore, 2006; Paris and Paris, 2001). The greater controllability of strategy attribution has significant educational implications. If success is explained in terms of effective learning strategies, then the student is more likely to sustain their use. Similarly, if failure is explained in terms of ineffective strategies, then the student is in a good position to improve learning by enhancing learning strategies (Chan and Moore, 2006).

5.3 *Avoiding common pitfalls in Assessment as Learning*

Several pitfalls should be avoided, and one is the assumption that self-directed learning can be developed as the child grows older. Research

(Desoete, 2008; Kramarski and Mevarech, 2003; Paris, 2002; Paris and Paris, 2001; Pintrich, 2002; Zimmerman, 2008) reveals that the opposite is true. These studies found that metacognitive knowledge and self-directed learning strategies must be taught explicitly, for instance by using explicit labeling, direct instruction, role modeling, open discussion, or reciprocal teaching. Children could not be expected to develop metacognition spontaneously as they mature and that children who had developed metacognitive skills might not know when to apply them. Embedding metacognitive instructions in the content matter and integrating the cognitive, motivational, and affective domains of self-directed learning provide the most effective approaches in developing in students sustainable self-directed learning competencies.

6 Conclusion

Fundamental changes have occurred in school assessment practices in Singapore, Hong Kong, and other parts of the world. Even though these countries are at different stages of economic development, all of them have initiated unprecedented reforms of their education and assessment systems in responding to the rapidly changing and globalizing international context.

The chapter draws on the work of the 3-year Assessment project (2005 to 2009) in Hong Kong, and the Self-directed Learning Orientated Assessment (SLOA) projects in Macau and China (2008-2009), some of the results have been reported in Mok (2010). This chapter presents the case of applying SLOA in mathematics education. The essence of SLOA is that assessment should serve and advance learning, and that the learner should be self-directed. This chapter discusses in detail the pedagogical approaches of the three integrated components, namely, Assessment *of* Learning, Assessment *for* Learning, and Assessment *as* Learning, tools for their implementation, and common pitfalls to avoid.

Valid assessment helps identify areas needing support, and with quality feedback derived from assessment, efficiencies of learning can be enhanced. SLOA is one model that can be applied to mathematics with proven success track records in China, Hong Kong and Macau.

Nevertheless, success with SLOA is not guaranteed. Critical success factors for the implementation of SLOA in mathematics are covered in the chapter.

Several tools have been introduced to facilitate the implementation of SLOA. Of these tools, cognitive diagnostic assessment and KIDMAP have great potentials for mathematics education.

References

Ballet, K., & Kelchtermans, G. (2009). Struggling with workload: Primary teachers' experience of intensification. *Teaching and Teacher Education, 25* (8), 1150-1157.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London: GL Assessment.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5-31.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning: Putting it into practice.* Berkshire, England: OUP.

Boekaerts, M., & Cascallar, E. (2006). How far have we moved toward the integration of theory and practice in self-regulation? *Educational and Psychological Review, 18*, 199-210.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*(2), 199-231.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Camahalan, F. M. G. (2006). Effects of self-regulated learning on mathematics achievement of selected southeast Asian children. *Journal of Instructional Psychology, 33*(3), 194-205.

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57-66.

Chacko, I. (1998). S-P chart and instructional decisions in the classroom. *International Journal of Mathematical Education in Science and Technology, 29*(3), 445-450.

Chan, W. Y., Lau, S., Nie, Y., Lim, S., & Hogan, D. (2008). Organizational and personal predictors of teacher commitment: The mediating role of teacher efficacy and identification with school. *American Educational Research Journal, 45*(3), 597-630.

Chan, L. K. S. & Moore, P. J. (2006). Development of attributional beliefs and strategic knowledge in years 5-9: A longitudinal analysis. *Educational Psychology, 26*(2), 161-185.

Day, C. (2008). Leadership and continuing professional development of teachers. In J. C. K. Lee & L. P. Shiu (Eds.), *Developing teachers and developing schools in changing contexts*. Hong Kong: The Chinese University Press.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics, 34*(1), 115-130.

Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: How you test is what you get. *Metacognition Learning, 3*, 189-206.

Desoete, A., Roeyers, H., & De Clercq, A. (2003). Can offline metacognition enhance mathematical problem solving? *Journal of Educational Psychology, 95*(1), 188-200.

Dewey, J., & Bento, J. (2009). Activating children's thinking skills. *British Journal of Educational Psychology, 79*(2), 329-351.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition Learning, 3,* 231-264.

Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.

Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. Weinert & R. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 21-29). Hillsdale, NJ: Erlbaum.

Harnisch, D. L. (1983). Item response patterns: Applications for educational practice. *Journal of Educational Measurement, 20*(2), 191-206.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Hogan, D., & Gopinathan, S. (2008). Knowledge management, sustainable innovation, and pre-service teacher education in Singapore. *Teachers and Teaching: Theory and Practice, 14*(4), 369-384.

Kan, F. (2010). The functions of Hong Kong's Chinese history, from colonialism to decolonization. *Journal of Curriculum Studies, 42*(2), 263-278.

Kayler, M.A. (2009). Teacher development and learner-centered theory. *Teacher Development, 13*(1), 57-69.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.).New York: Springer Science and Business Media.

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal 40*(1), 281-310.

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition Learning, 5*, 173-194.

Lau, D. C. H., Mok, M. M. C. & Yan, Z. (2009, Nov). *An assessment scale for measuring primary and secondary mathematics achievement*. Paper presented at the International Conference on Primary Education 2009, 25-27 November 2009, The Hong Kong Institute of Education.

Law, Y. K., Chan, C. K. K., & Sachs, J. (2008). Beliefs about learning, self-regulated strategies and text comprehension among Chinese children. *British Journal of Educational Psychology, 78*, 51-73.

Lee, J. C. K., Yin, H., & Zhang, Z. (2009). Exploring the influence of the classroom environment on students' motivation and self-regulated learning in Hong Kong. *The Asia-Pacific Education Researcher, 18*(2), 219-232

Mok, M. C. M. (2010). *Self-directed learning oriented assessment: Assessment that informs learning and empowers the learner*. Hong Kong: Pace Publishing Ltd.

Ministry of Education. (2010). *Desired outcomes of education*. Retrieved August 7, 2010, from http://www.moe.gov.sg/education/desired-outcomes/

OECD. (2004). *Lifelong learning: OECD Policy brief*. Retrieved August 7, 2010, from http://www.oecd.org/dataoecd/17/11/29478789.pdf

Paris, A. H., & Paris, S. G. (2001). Classroom application of research on self-regulated learning. *Educational Psychologist, 36*(2), 89-101.

Paris, S. G. (2002). When is metacognition helpful, debilitating, or benign? In P. Chambres, M. Izaute, & P. J. Marescaux (Eds.), *Metacognition: Process, function and use* (pp. 105-120). Boston: Kluwer Academic Publishers.

Paris, S. G., & Winograd, P.W. (1990). How metacognition can promote academic learning and instruction. In B.J. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp.15-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Patz, R. J., & Yao, L. (2007). Vertical scaling: Statistical models for measuring growth and achievement. In C.R. Rao, & S. Sinharay (Eds.), *Psychometrics* (pp. 955-978). Amsterdam: Elsevier North-Holland.

Phillipson, S. (2006). Cultural variability in parent and child achievement attributions: A study from Hong Kong. *Educational Psychology, 26* (5), 625-642.

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice, 41*(4), 219-225.

Rieber, R. W., & Carton, A. S. (1987). *The collected works of L.S. Vygotsky.* Translated and with an introduction by Norris Minick. New York: Plenum Press.

Sato, T. (1980). The S-P Chart and the caution index. Tokyo, Japan: *NEC Educational Information Bulletin*, 80-1, C&C systems Research Laboratories, Nippon Electric Co., Ltd.

Schunk, D. H. (1998). Teaching elementary students to self-regulate practice of mathematical skills with modeling. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning and performance*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Schunk, D. H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review, 20,* 463-467.

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly, 23*, 7-25.

Spektor-Levy, O., Eylon, B. S., & Scherz, Z. (2008). Teaching scientific communication skills in science: Tracing teacher change. *Teaching and Teacher Education, 24*, 462-477.

Stoeger, H., & Ziegler, A. (2008). Evaluation of a classroom based training to improve self-regulation in time management tasks during homework activities with fourth graders. *Metacognition Learning, 3*, 207-230.

Swartz, R., & Parks, S. (1994). *Infusing the teaching of critical and creative thinking into content instruction: A lesson design handbook for the elementary grades*. Pacific Grove, CA: Critical Thinking Press and Software.

Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to rule space method.* New York: Routledge Taylor & Francis Group.

Vermeer, H., Boekaerts, M., & Seegers, G. (2001). Motivational and gender differences: Sixth-grade students' mathematical problem-solving behavior. *Journal of Educational Psychology, 92*(2), 308-315.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.

Winne, P. H. (2005). A perspective on state-of-the-art research on self-regulated learning. *Instructional Science, 33*, 559-565.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13 – 39). San Diego, CA: Academic Press.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183.

Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. *Journal of Educational Psychology, 91*(2), 241-250.

Chapter 10

Building Bridges Between Large-Scale External Assessment and Mathematics Classrooms: A Japanese Perspective

Yoshinori SHIMIZU

This chapter discusses how large-scale external assessment and classroom assessment in mathematics can be linked for enhancing students' learning and improving classroom teaching with a particular reference to the case of the new National Assessment of Academic Ability in Japan. The framework for mathematics assessment and sample items is provided to describe how the test items and the assessment results can be used for the improvement of classroom teaching. There is a tension between large-scale assessment and classroom assessment in their differences of purpose, method, emphasis, and audience. Large-scale external assessment, however, needs not be seen as completely different from classroom assessment. Rather, external assessment like the one discussed in this chapter can be used in certain ways for enhancing students' learning and for improving classroom teaching.

1 Introduction

Assessment has been a long-standing problem in Japanese mathematics education, as it has been in Singapore. Japanese mathematics educators have struggled for decades with many of the same assessment issues that plague educators in Singapore, asking questions such as:

- What influences, both positive and negative, are exerted by external assessment on classroom-based assessment?
- How can the results of external assessments be used to design activities to move students' thinking forward, in addition to providing evidence of their present levels of knowledge and skill?
- How can teachers become more familiar, through assessment, with the abilities, skills, and thinking of their students, and thereby more appropriately able to plan and modify their instruction?

In April 2007, for the first time in 43 years, Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) conducted its national assessment of academic ability in school subjects of Japanese and Mathematics. The assessment aims to monitor student's academic ability and backgrounds of their learning nationwide and to examine and improve educational policies, and to provide key information to local boards of education and schools so that they can improve classroom practices. The new external assessment, started with the entire cohorts in grades 6 and 9, had strong impacts on classroom practices in those subjects.

In any school subject, in general, and in mathematics, in particular, a tension between large-scale external assessment and classroom assessment exists in their differences of purpose, method, emphasis, and audience. One of the key, but sometimes not noted, issues for classroom teachers with large-scale external assessments is to think about how to capitalize the released results of them for improving classroom practices.

In the following sections, I will discuss how large-scale external assessment and classroom assessment in mathematics can be linked to enhance students' learning with a particular reference to the case of the new national assessemnt of academic ability in Japan. After the Japanese contexts of the introduction of the national assessment are described briefly, the framework for the new mathematics assessment and several sample items are provided to describe how the test items and the results can be used for the improvement of classroom practices. It is then argued that large-scale assessment needs not be seen as completely different from classroom assessment and that external assessment like the one discussed in this chapter can be used in certain ways to enhance students' learning.

2 Recent National Assessments in Japan

2.1 *The Japanese contexts*

The Japanese education system is comprised of 6 years of elementary school, 3 years of lower secondary school, 3 years of upper secondary school, and 2-4 years of postsecondary school (e.g., 2 years of junior college or 4 years of university). Recently, secondary schools of 6 years are also available.

The basic guidelines for school curricula to be used nationwide are prescribed in the National Course of Study, which is issued by the MEXT. The document includes the objectives and contents of all the school subjects. Each school sets up and implements its own curricula in accordance with the guidelines, taking account the conditions of the local community and the school, the stages of growth and the characters of students, as well as other conditions for students' learning.

The MEXT has recently introduced a national assessment at the final grades, the "exit", of elementary and lower secondary schools to monitor student's academic ability and backgrounds of students' learning nationwide. In the current Japanese system, students' learning is evaluated by criterion-oriented evaluation from four different viewpoints. For mathematics, evaluation of students' learning covers the following four categories.

- Interests in, eagerness for, and attitudes toward mathematics
- Mathematical ways of thinking
- Ability to represent and process mathematical objects
- Mathematical knowledge and understanding

The new national assessment discussed in this chapter aims to assess the last three categories by paper and pencil tests and the first category by questionnaires.

2.2 *Large-scale assessment in Japan*

Besides the large-scale international assessments such as IEA's TIMSS (*Trends in International Mathematics and Science Study*, see Mullis et al., 2004) and OECD's PISA (*Programme for International Students Assessment*, see OECD, 2004), several types of large-scale assessments, including paper-and-pencil tests with questionnaires administered to students, teachers, and schools, have been conducted in Japan. In particular, since the 1980s, three different types of large-scale assessments have been implemented; the Assessment of Implementation of National Curriculum, the Assessment of Specific Issues in Students' Learning, and the National Assessment of Academic Ability and Learning Environments. Each of these assessments has different aims and objectives for different school subjects with different student groups as shown in Table 1. Assessment of Implementation of National Curriculum has been conducted roughly every ten years to monitor the implementation of new National Course of Study and then to improve classroom practices in each school (National Institute for Educational Policy Research, 2005). Assessment of Specific Issues in Students' Learning plays a complementary role to it by scrutinizing students' difficulties identified by other assessments.

Table1

Types of recent large-scale national assessments in Japan

	National Assessment of Academic Ability and Learning Environments	Assessment of Implementation of National Curriculum	Assessment of Specific Issues in Students' Learning
Major Aims	To monitor student's academic ability and background of learning nationwide to check and improve educational policy. To establish the PDCA (Plan-Do-Check-Action) cycle in educational policy. To improve classroom practices in each school.	To monitor the implementation of new national course of study. To improve classroom practices in each school.	To investigate specific issues in teaching and learning which are not explored by the Assessment of Implementation of National Curriculum.
Targeted Grades	Grades 6 and 9	Grades 5 through 9	Depends on the subject (Grades 4 through 9 for Mathematics)
Survey Style	Complete (2007-2009) Sampling (2010-)	Sampling	Sampling
School Subjects	Japanese and Mathematics	Japanese, Mathematics, Social Studies, Science, and English (only for junior high schools)	All the school subjects

3 The New National Assessment of Academic Ability and Learning Environments

In April 2007, the new nationwide test was implemented to assess the academic achievement of sixth-graders in elementary schools and third-graders in junior high schools as well as to investigate the environments and situations of students' learning in and out schools. Students in the targeted grades are in their final years of schooling at each stage. Their scores in the test can be considered to give a good indication of how much progress they have made at those stages of their education.

From 1956 to 1966, there were national achievement tests covering random sample (5-10%) of all the students, and another test for all students in grades 8 and 9. However, these tests were suspended as they seemed to accelerate competitions among schools. The new nationwide test in 2007 was a response to public concerns over the deterioration in academic skills that became evident since 2002, when textbook content was reduced by roughly 30%, the school week was shortened from 6 days to 5, and the Japanese ranking "went down" from PISA 2000 to PISA 2003. In PISA 2000, Japan was on the top of the list of countries and regions in terms of students' achievement in mathematics, but in PISA2003, Japan was in the sixth place.

3.1 *The framework for mathematics assessment*

The new national assessment consists of two bundles, A and B, for both Japanese and Mathematics. Each of two bundles covers "Knowledge" and "Functional Use" respectively in each subject as described in the followings.

- Bundle A, Items for assessing "Knowledge": Knowledge and skills needed for further learning in schools and for applying in the real life situations
- Bundle B, Items for assessing "Functional Use": Competencies for applying knowledge and skills to the situations in the real life, and for planning, implementing, reflecting, and improving the plan to solve problems

Students in grade 9 worked on each bundle for 45 minutes, followed by another 45 minutes for the questionnaire. Bundle A includes multiple choice and short answer style tasks, while bundle B includes open construction tasks as well.

This is a summative assessment in nature based on the current curriculum. The results give the ministry vital information on students' academic performance nationwide, which will also be provided to schools and students. From 2007 to 2009, about 1.2 million sixth-graders at 22,000 elementary schools and 1.2 million third-year students at 10,500 junior high schools took the test. From 2010, only a random sample of the targeted students took the test.

Every year after a couple of months after its implementation, the MEXT releases the results of assessment to the local governments, boards of education, and schools that participated in the assessment. Also, the students who participated in the test obtain feedback on their papers and other information, including charts showing statistical information on the test. Finally, classroom teachers are provided with documents that describe detailed information of the intention of items and results from related items in the previous assessment, as well as recommended lesson plans so that they can use the assessment tasks in their classrooms.

3.2 *The framework for mathematics bundle B*

While each item in bundle A is intended to assess students' basic knowledge and skills, the items in the bundle B are to assess students' functional use of mathematics in various contexts such as daily lives, learning in other school subjects like science and social studies, and learning within mathematics. Key phrases that describe the abilities the assessemnt tasks require of the students in bundle B for grade 9 are as follows (National Institute for Educational Policy Research, 2008).

- Observing things around us by focusing on numbers, quantities, and figures to grasp their key features
- Classifying the given infromation to select an appropriate one
- Thinking logically to draw a conclusion and looking back on one's own thinking
- Interpreting events in the real world or in mathematical world, and expressing ideas mathemtically

There are three dimensions to the mathematics assessment tasks in bundle B:

- Mathematics content specified in the National Course of Study
- Situations and contexts
- Mathematical processes

First, each item is aligned with the National Course of Study that specifies goals and content of school mathematics. Second, there are three categories for situations and contexts with which students are faced in the test: mathematics, other school subjects, and the real world. Third, there are three strands α , β , and γ in mathematical processes:

α : Competencies for applying knowledge and skills to the situations in the real life

β : Competencies for planning, implementing, evaluating, and improving the plan to solve problems

γ : Related to both α and β

Table 2 shows the detailed description of mathematical processes to be assessed.

Table 2

The mathematical processes to be assessed by the tasks in bundle B

Category	The mathematical processes
Competencies for applying knowledge and skills to the situations in real life	α 1: Mathematizing phenomena in everdy life. α 1(1) Observing things by focusing on numbers, quantities, and shapes α 1(2) Grasping key features of things around us α 1(3) Idealizing and simplifying α 2: Functional use of information α 2(1) Classifying and organizing the given infromation α 2(2) Sellect needed information appropriately to make decisions α 3: Interpreting and expressing phenomena mathematically α 3(1) Interpreting phenomena α 3(2) Expressing own idea mathematically
Competencies for planning, implementing, evaluating, and improving the plan to solve problems	β 1: Drawing up a plan to solve problems β 1(1) Thinking in a logical way β 1(2) Making a plan β 1(3) Implementing the plan β 2: Evaluating the result and improving the entire process β 2(1) Looking back the result β 2(2) Improving the result β 2(3) Extending the result
Related to both α and β	γ 1: Making connections from one phenomenon to the other γ 2: Integrating different things γ 3: Considering things from multiple perspectives

Each item in bundle B is developed within the framework and described based on the three dimensions. The key feature of items in bundle B, as embedded in the real world contexts with an emphasis on mathematical processes, is new to the teachers.

3.3 *Emphasis on explaining mathematically*

The new National Course of Study emphases the importance of "language activity" in classroom to be facilitated in each subject to enhance students' learning. In accordance with the emphasis in the revised curriculum guidelines, each of the tasks in bundle B includes open-construction tasks that requires students to explain things in one of the following forms of explanation:

- The observed facts and properties in a situation
- Approaches and methods for solving a problem
- Reasons for the facts and properties

The first category corresponds to tasks that ask students to describe a mathematical fact or property, mostly in the form of a proposition. Tasks correspond to the second category ask students to describe the approach to a problem by specifying both "what is used" and "how it is used" as described in the example of "Mt. Fuji" shown below. It should be noted here that an "answer" is not required and the focus is on the method to be used. The third category includes tasks that ask students to explain the reason for the facts and properties. Construction of a proof falls into this category.

3.4 *Sample items and the results*

The results of assessment were released a couple of months after its implementation, on July 31st in the case of year 2010. For grade 9 in 2010, the average percentage of correct response to items in bundle A ("Knowledge") was 66.1, while for bundle B ("Functional use") it was 45.2. In general, while the percentages of correct responses were high for the knowledge and skill items, there were several items that show the students' difficulties in mathematics. Also, the percentages of correct response to items in bundle B were relatively low. Some sample items are shown below.

1. Common and decimal and fractions (Grade 6, A$\boxed{3}$ (2), 2007)

Choose the largest number among 0.5, 7/10 and 4/5, and then show the locations of the numbers on the number line.

The percentage of students who answered the correct choice was 55.9. About 17% of the students chose "7/10" as the largest number. The result reveals that students have difficulties in understanding and expressing the size and meaning of common and decimal fractions.

2. The relationship between divisor and quotient (Grade 6, A$\boxed{3}$, 2008)

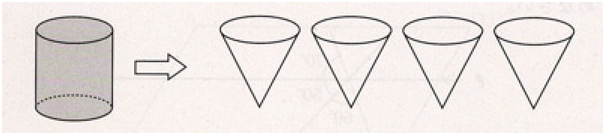
In the following expressions, "@" denotes the same number which is not 0. Choose all the expressions that the result of computation is larger than the original number @.

a. @ × 1.2
b. @ × 0.7
c. @ ÷ 1.3
d. @ ÷ 0.8

The percentage of students who answered the correct choices (a and d) was 45.3. Those who chose "a" and "c" (12.0%) thought that the number in the expression, being larger than 1, will make the result of computation larger than "@". The small group of students (4.4%) who chose "a" and "b" might think that multiplication makes the result of computation larger than the original number.

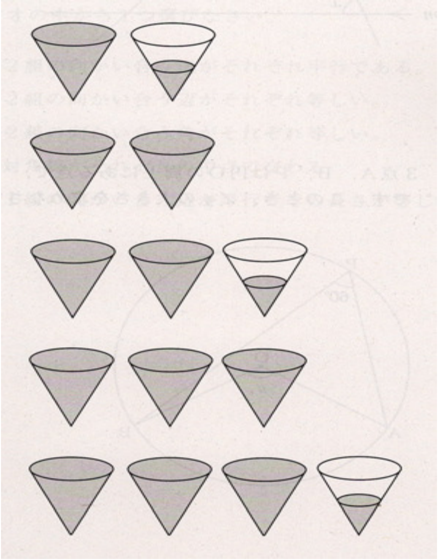
3. Volumes of cylinder and cone (Grade 9, A$\boxed{5}$ (4), 2007)

The following figures represent containers of cylinder and cone with the same height and the diameter of the circular top.



When we move the water in the cylinder to the cone, how many cones are filled by water? Choose the correct figure from the following five choices.

a.
b.
c.
d.
e.



Only 38.1% of the students chose the correct answer (d). Roughly the same percentage of students (36.7%) chose wrong choice “b”. The result clearly shows that students’ understanding of the relationship between the ratios of volumes of cylinder and cone is very weak and that this may be caused by the lack of students’ experience to have the activity in mathematics classrooms.

4. Two-digit numbers (Grade 9, A$\boxed{2}$ (4), 2010)

When we express a two-digit number by using x for the tenth digit and y for the unit, which of the following expressions is correct? a. xy b. $x + y$ c. $10\,xy$ d. $10x + y$

The result of the item is shown in Table 3. Many students had difficulties in representing a two-digit number using letters; they could not correctly make connections between the ten base numeral system with literal symbols.

Table 3

Results of the Item A$\boxed{2}$(4), Grade 9, 2010

Category	Choice	Response rate (%)
1	xy	11.5
2	$x + y$	11.0
3	$10xy$	8.9
4	$10x + y$	*67.7
9	Others	0.2
0	No Answer	0.8

5. "Mt. Fuji", An Item from Bundle B (Grade 9, B$\boxed{5}$, 2008)

This item (see Appendix) is intended to assess students' ability to apply linear function to real data to solve a problem in context. It involves using data in a given table and a graph of linear function to interpret phenomenon in the real world. Students need to use mathematics and explain a method for estimating the temperature at a certain location.

In this situation, air temperature (y) is treated as a linear function of altitude (x), and students are asked to explain a method for finding air temperature at the altitude of 2500m by describing both "what is used" and "how it is used". Here the category of "what is used" includes graphs, expressions, tables, numerical values, and so on. The category of "how it is used", on the other hand, includes drawing a straight line to identify the value of y when $x = 2500$, finding the expression of linear function from the data given, and examining the rate of change from the table, and so on.

The correct response rates were, 54.7%, 25.0%, and 13.3% for Question 1, 2, and 3, respectively. The rates were quite low for Questions 2 and 3. Most students had difficulties in finding and applying linear function in the given context. Also explaining their approach to the problem by specifying both "what is used" and "how it is used" was difficult for them, when the "answer" was not requested. The no response rate for Question 3 was 58.5%, the highest in all the items in the year.

4 Linking Large-Scale External Assessment to Classroom Practice

As mentioned earlier, there is a tension between large-scale assessment and classroom assessment. Classroom assessment is designed or used by classroom teachers for making instructional decisions, monitoring students' progress, or evaluating students' achievement (Wilson and Kenney, 2003). On the other hand, large-scale assessments are summative in nature and external to the course of regular classroom activities. Nevertheless, even in the case of large-scale external assessment, "assessment should enhance mathematics learning" (NCET, 1995, p.13). Given the fact that taking a test, even an external one, is a key learning opportunity for students, assessment tasks in the external test need to be considered as the platform for enhancing students' learning. Also, it should be noted that the strong and close relationship between assessment and instruction has great potentials for improving classroom practice (Van den Heuvel-Panhuizen and Becker, 2003). How can we make a link between large-scale external assessment and classroom instruction in mathematics?

4.1 *Using external assessment task to improve classroom practices*

In the context of the current educational reform in Japan, the relationship between the revised national curriculum framework and the national assessment is a key to the implementation of new curriculum. If the emphasis in the new curriculum is reflected in the new national assessment, the test items play a key role in sending messages on the emphasis in the new curriculum standards to classroom mathematics teachers, just as "what is tested is what gets taught". In the case of the new national assessment in Japan, inclusion of open-constructed items with an emphasis on "explaining" is aligned with the emphasis on "language activity" in the new National Course of Study.

While the new national assessment focuses on fundamentals in school mathematics, the test items, those in bundle B, in particular, are new to most teachers, as they are embedded in the real world or intra-mathematics contexts with an emphasis on mathematical processes that are key features of the new National Course of Study. Teachers can become more familiar, through the new types of assessment tasks, with abilities, skills, and thinking of their students that need to be fostered, and thereby more appropriately able to plan and modify their instruction.

The MEXT has started to disseminate leaflets that include both the results on a few items from the assessment and recommended lesson plans related to those items. The leaflets are sent to all the schools to promote new visions among teachers on classroom teaching with the new emphasis. In this sense, external assessments can be used to design lessons to move students' thinking forward, in addition to providing evidence of their present levels of knowledge and skill.

4.2 *Using the results of external assessment to help students*

The assessment result of each item can be used to inform classroom teaching with more attention paid to those students with difficulties related to the item. This is the case when an assessment task is aligned with the mathematics content specified in the National Course of Study. In other words, test items can be connected to students' learning, even though they are external to the course of regular classroom activities, in terms of mathematics content in the curriculum.

As for the example of how teachers can help their students by using the assessment results, look at Table 3. Choice a, "*xy*", corresponds to those students who just replaced the numbers in each digit with letters x and y. Choice b, "$x + y$", corresponds to those students who did not understand the magnitude of the number in the tenth digit. Finally, choice c, "$10xy$", corresponds to those students who understood the magnitude of the number in the tenth digit but could not represent by using literal symbols. Assessment result suggests how the teacher can help the students. If you have the students whose tend to choose "b" in your classroom, for example, you may show to the students a two digit number, say, 24, and make sure that the sum of x and y equals 6, not 24.

The MEXT has provided classroom teachers with suggestions and implication on what they can do with the students with difficulties as specified with the coding for analysis for all the items (e.g. National Institute for Educational Policy Research, 2010). As the "Learning" standard from the NCTM Assessment Standards (NCTM, 1995) suggests, the assessment task provides "learning opportunities as well as opportunities for students to demonstrate what they know and can do" (p.13). From the teachers' perspective, the results of each item in large-scale assessments provide information about expected students' difficulties.

5 Concluding Remarks

Classroom teachers often do not have direct influence on external assessment programs but such programs do have significant influence on what happens in classrooms. Although there is a tension between large-scale assessment and classroom assessment, as mentioned earlier, large-scale assessment needs not be seen as completely different from classroom assessment. Rather, external assessment, such as the one discussed in this chapter, fits with classroom assessment and results can be used for anticipating and considering students' thinking in each content domain.

Assessment should be aligned with and central to teaching mathematics. Students nowadays are living in the era of external assessment. Given the fact that taking a test is also a key learning opportunity for students, assessment tasks in an external assessment can be a platform for enhancing students' learning and for improving classroom teaching.

References

Ministry of Education, Culture, Sports, Science, and Technology (2009). *The national course of study*. Tokyo: Author.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eight grades.* Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

National Council of Teachers of Mathematics (1995). *Assessment standards for school mathematics.* Reston, VA: Author.

National Institute for Educational Policy Research (2005). *A summary of 2003 national survey of the implementation of the curriculum: Mathematics*. Tokyo: Author.

National Institute for Educational Policy Research (2008). *Results from 2008 national assessment of students' academic achievements and learning environments*. Tokyo: Author.

National Institute for Educational Policy Research (2010). *Results from 2010 national assessment of students' academic achievements and learning environments*. Tokyo: Author.

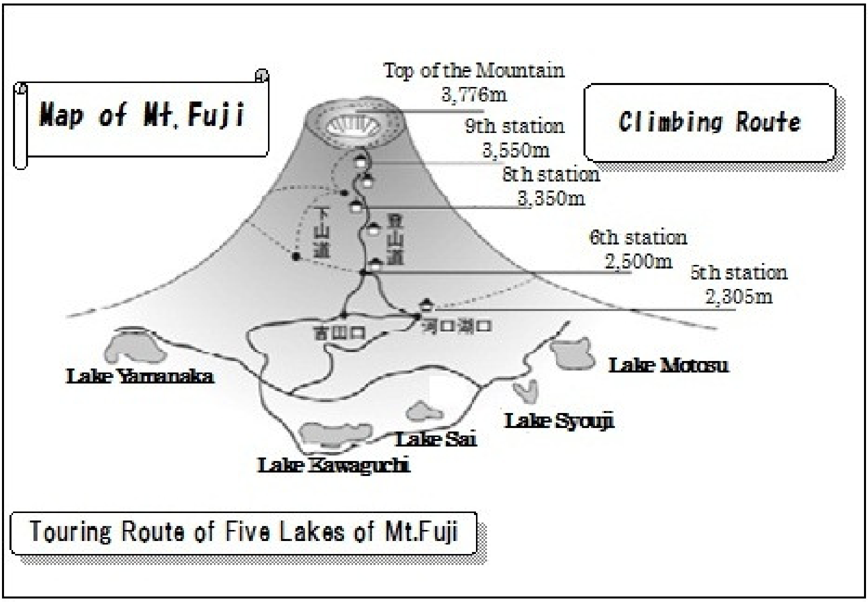
Organisation for Economic Co-operation and Development (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.

Van den Heuvel-Panhuizen. M., & Becker, J. P. (2003). Toward a didactic model for assessment design in mathematics education. In A.J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick & F.K S. Leung (Eds.), *Second international handbook of mathematics teaching and learning* (pp. 689-716). Dordrecht: Kluwer Academic Publishers.

Wilson, L. D., & Kenney, P. A. (2003). Classroom and large-scale assessment. In J. Kilpatrick, W. G. Martin & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 53-67). Reston, VA: National Council of Teachers of Mathematics.

Appendix

Rina and her friends are planning to visit the Five Lakes of Mt. Fuji and then climb up to the sixth stage of the mountain this August.



A Map of Mt. Fuji Climbing and the Five Lakes of Mt. Fuji

Question 1.
You will take photos at two lakes among the five. How many different choices of two lakes do you have, if we ignore the order of the visits?

Question 2.
Rina and Ken-ichi are talking about the temperature of the sixth stage of Mt. Fuji.

Rina: I have tried to investigate the temperature of the sixth stage, but I couldn't find it because there is no observatory on the stage.
Ken-ichi: <u>It is known that the temperature falls at a constant rate as one climbs higher until an altitude of 10,000 meters</u>.
Rina: We may use the fact to find the temperature of the sixth stage.

If we hypothesize that the temperature falls at a constant rate as one climbs higher until an altitude of 10,000 meters, what is the relationship that holds anytime between altitude x meters and temperature y °C? Choose the correct one from the followings.

a. y is proportional to x.
b. y is an inverse proportion to x.
c. y is a linear function of x.
d. Sum of x and y is a constant.
e. Difference of x and y is a constant.

Question 3.
Rina investigated the mean temperature in August on the top of the mountain and around Mt. Fuji. She completed table below and drew a graph, measuring altitude as x meters and temperature y °C.

Table 1
Altitude and Mean Temperature in August at Observation Points

Observation Points	Altitude (m)	Mean Temp (°C)	Observation Points	Altitude (m)	Mean Temp (°C)
A (Kofu)	273	27.7	D (Kawaguchiko)	860	23.3
B (Katsunuma)	394	26.7	E (Yamanaka)	992	21.7
C (Furuseki)	552	24.9	F (Fujisan)	3775	6.4

(*Data Source: Meteorological Agency*)

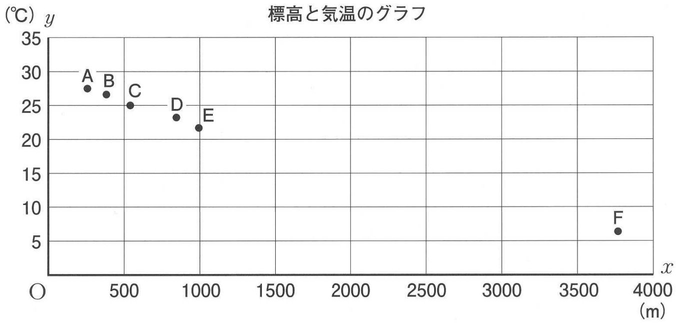


Figure 1. Graph of the relation between Altitude and Temperature

Rina understood that the temperature falls at a constant rate as one climbs higher. Then, she tried to estimate the temperature of the sixth stage of Mt. Fuji using data at the points D and F in the given table and the graph. Explain your method of estimating the temperature at the sixth stage (2,500m). You do not need to actually find the temperature.

Chapter 11

Errors in Mathematics Assessment Items Written by Pre-Service Teachers

Jaguthsing DINDYAL

Learning to write good test items is an important aspect of the teacher preparation programmes in Singapore. This chapter highlights the types of errors in mathematics test items made by student teachers who were following their pre-service course for teaching at primary level. An analysis of the errors reveals that these student teachers demonstrate some key shortcomings when writing test items: use of language, mastery of content knowledge, use of diagrams as scaffolds, and the use of appropriate context.

1 Introduction

It is often claimed that assessment drives the curriculum. The National Research Council (1989) in the United States acknowledged this fact by stating that "what is tested is what gets taught" (p. 69). It is not surprising that teachers who implement the school curriculum are often teaching to the test. Traditional paper and pencil tests are still the norm for assessing students' learning in schools. Thus, assessment is high on the agenda of any school mathematics teacher, in an examination-oriented system such as Singapore. They have to look very carefully into both formative and summative aspects of assessment as parents and the community at large, are very sensitive to any form of assessment in which students are involved. Any shortcoming in assessment practices can entail serious consequences for teachers and their schools.

School assessments are only as good as the individual test items that make up these assessments. When teaching mathematics, teachers use a large number of problems as test items. Some of the items are copied from textbooks and past examination papers, whereas others are modified from similar sources to match the specific nature of their classes. In addition, a fairly large number of items for tests are originally written by the teachers themselves. Hence, it is imperative that mathematics teachers take extreme care in writing their test items. In this chapter, I report on the types of errors that were noted when a group of pre-service primary teachers wrote test items. Four categories of errors were identified and these are described in detail in the latter part of the chapter.

2 Mathematical Tasks

Lester (1983) referred to mathematical problems as tasks. Good tasks are the ones that do not separate mathematical thinking from mathematical concepts or skills, they capture students' curiosity, and they invite the students to speculate and to pursue their hunches (National Council of Teachers of Mathematics [NCTM], 1991). The typical test includes several items or problems. A problem can be considered as a *task* which elicits some activity on the part of students and through which they learn mathematics during the problem solving activity. Several factors can be identified that differentiate one problem from another. Amongst others, problems differ by: the content domain, the objectives to be tested, the exact wording of the problem, the context of the problem, the support and structure provided, the types of numbers involved, the resources to be used during the solution process, the expected time for a solution, and the *closedness* or *openness* of the problem (see Dindyal, 2009).

Mason and Johnston-Wilder (2006) highlighted the various ways in which a mathematical task is perceived: (1) the task as imagined by the task author, (2) the task as intended by the teacher, (3) the task as specified by the teacher-author instructions, (4) the task as construed by the learners, and (5) the task as carried out by the learners. It is to be noted that whether a problem task is originally written down by a teacher or is taken or modified from a secondary source, it carries an implicit intent that the assigner of the task (in this case the teacher) wishes to achieve by assigning the problem task to the solver (in this case the student). There are bound to be mismatches between what the assigner wishes to achieve and what actually is achieved during the solving process. Accordingly, we must aim to reduce if not eliminate these mismatches. One of the ways to reduce the mismatches is to produce items that would be construed and carried out by the learners in a way that was initially intended. This implies writing good assessment items.

Some of the issues we have with assessment items involve deficiencies in the technical aspects of writing the problems. Unless an assessment item is written carefully, there are bound to be inconsistencies in the way these items are perceived by the students who work on them in their tests. Noddings (1988) has claimed that structural features in the problems themselves account for some student difficulty.

Verschaffel, Greer, and De Corte (2000, p. x), who were referring to word problems, claimed that several components can be distinguished in a word problem:

- The mathematical structure: i.e., the nature of the given and unknown quantities involved in the problem, as well as the kind of mathematical operation(s) by which the unknown quantities can be derived from the givens.
- The semantic structure: i.e., the way in which an interpretation of the text points to particular mathematical relationships – for example, when the text implies a change from an initial quantity to a subsequent quantity by addition or subtraction, or a combination of disjoint subsets into a superset, or the additive comparison between two collections, then in each case operations of addition or subtraction are indicated.
- The context: What is the problem about, e.g., whether, in the case of an additive problem involving combination of disjoint sets, it deals with groups of people joining each other, with collections of objects, etc.
- The format: i.e., how the problem is formulated and presented, involving such factors as the placement of the question, the

complexity of the lexical and grammatical structures, the presence of superfluous information, etc.

The mathematical structure, the semantic structure, the context and the format can provide a framework to look at particular problems that primary students solve. Verschaffel, Greer, and De Corte's (2000) characterization of word problems is quite similar to Kulm's (1979) categories of task variables. Kulm commented that task variables can be broadly categorized as those that: (1) describe the problem syntax, (2) characterize the problem's mathematical content and non-mathematical context, (3) describe the structure of the problem, and (4) characterize the heuristic process evoked by the problem. Although word problems at the primary level are traditionally discussed within the domain of numbers, the points mentioned above are generally applicable to problems from other domains. The framework can also provide a lens for looking at deficiencies in items. A poorly written test item can be defective in one or more of the above categories.

3 Errors in Mathematics Assessment Items

The errors in the mathematics assessment items that are described in this chapter were identified in the test construction task which is a major assignment of pre-service teachers following the final year of the degree programme for teaching at the primary level. In their first year of study, these pre-service teachers have covered two content courses: one on number and algebra and the other one on geometry and data. In each of their next three years of study, the student teachers take a course on the teaching and learning of mathematics at the primary level. The last of these courses in their fourth year of study is split over two semesters, of which in the last semester the students take a course on assessment in mathematics and the planning and construction of test items for mathematics tests and examinations. Student teachers learn how to construct a table of specifications (TOS) for constructing a test and how to develop test items based on the TOS. Issues about test objectives, types of items, difficulty level of items, types of marks and marking schemes are discussed during this course. They are also exposed to alternative modes of assessment. The items discussed here were collected from the test that the degree pre-service teachers developed after following the course as part of their final project. As a marking scheme had to be provided, the pre-service teachers' solutions were helpful in categorizing the items and providing a window on their thinking when they developed the items. Often, there were mismatches between what these pre-service teachers wrote as items and what they actually solved. An analysis of the test items, together with the proposed solutions, was helpful in categorizing the types of errors in the test items. The errors in the mathematics test items are divided into four groups: errors in the use of language, errors due to a poor mastery of content knowledge, errors due to poor diagrams as support, and errors due to an inappropriate context.

3.1 *Language-related errors*

It was noted that errors in assessment items due to the language were of several types:

1. *Unclear instructions*
These types of errors do not give a clear indication about what to do in the item. For example, in the item described in Figure 1, the writer was not clear about finding the sum of 9.7 and 0.5 and then dividing by 100. This type of errors can be generally attributed to a lack of mastery of the language of communication. Hence, it leads to a mismatch between what the writer intends and what is actually written down.

8. What is the sum of 9.7 and 0.5 when divided by 100?

Ans: __________

Figure 1. Language errors: unclear instructions

2. *Missing key word or phrase*

This type of error occurs when certain key words or phrases are omitted from the instructions in the item. For example, in the item described in Figure 2, the key word "equal" was omitted which makes it impossible to solve as the problem was not meant to be open ended. Unless it is clearly stated that there are 8 equal parts, the problem cannot be solved.

8. Jie Jern cuts a 16.8 m of string into 8 parts. What is the length of each piece of string?

Figure 2. Language errors: missing key word or phrase

3. *Using the incorrect direction verb*

This type of error occurs when the incorrect direction verb (words such as find, calculate, show, estimate, draw, prove, etc.) is used. In the item described in Figure 3, it does not make sense to say "prove your answer clearly". For students at the primary level this term can be quite confusing.

Alex and Peter are asked to arrange the following decimals in increasing order.

0.612 , 0.81 , 0.062 , 8.12 , 6.21

The table below shows the responses of Alex and John respectively.

Alex	0.062 , 0.81 , 0.612 , 6.21 , 8.12
Peter	0.062 , 0.612 , 0.81 , 6.21 , 8.12

Which boy has answered wrongly? Prove your answer clearly.

Figure 3. Language errors: incorrect direction verb

4. *Incorrect description of the context*

This type of error is due to a poor description of the context. Usually it is also due to a poor grasp of the language of communication. In the item in Figure 4, the sentence "You have a clock that is spoilt from 12 pm onwards", very poorly describes what the writer of the item actually meant.

> 3)
> You have a clock at home that is spoilt from 12pm onwards, for every 5 minutes that it moves, the minute hand will slide back 1 minute. If the clock shows 5pm, what is the actual time?

Figure 4. Language errors: incorrect description of the context

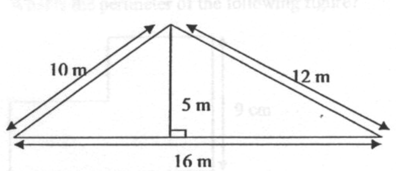
3.2 *Content-related errors*

Some of the errors in the assessment items are identified as being due to a lack of content knowledge.

1. *Over-defined conditions*

These types of errors are noted particularly in geometry items. Typically the writers of these items disregard some of the constraints under which a geometrical figure may or may not exist. One such item is shown in Figure 5, in which the lengths of the sides of the largest triangle clearly determine the height of the triangle. By giving the height of the triangle as 5 m, the writer ignores this fundamental geometrical property.

> 12. The figure shows a triangle. Find the area of the triangle.

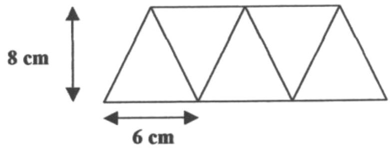


> ______________ m^2

Figure 5. Over-defined conditions 1

In the item in Figure 6, a similar situation arises. The sides of the equilateral triangle amply define the height of the triangle. By giving an arbitrary value to this height, the writer of this item provided an artificial context in which many students would be tempted to give the area of one of the equilateral triangles as "half of the base times height".

2.

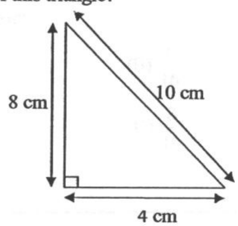


The figure above is made up of 5 equilateral triangles. Find its total area.

Figure 6. Over-defined conditions 2

For the item in Figure 7, there was no requirement for the triangle to be right-angled if the skill that the writer of the item was looking for was to find the perimeter when three sides of a triangle were given. However, by showing the triangle to be right-angled, the writer overlooked the fact that the conditions no longer satisfy Pythagoras theorem.

7. What is the perimeter of this triangle?



1) 16 cm
2) 20 cm
3) 22 cm
4) 40 cm

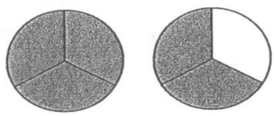
( )

Figure 7. Over-defined conditions

2. *Mathematical concepts*

This type of errors are noticed when the writers demonstrate a lack of knowledge about certain basic mathematical concepts. For example, in the item in Figure 8, the writer is not clear about what the *figure* was or what the whole is. Is one circular shape the required figure or do the two circular shapes together constitute the figure? As the two possible responses can be found in the options for the answers, this multiple choice item is misleading.

What fraction of the figure below is shaded?



(1) $\frac{5}{6}$

(2) $\frac{3}{5}$

(3) $\frac{6}{5}$

(4) $\frac{5}{3}$ ( )

Figure 8. Mathematical concepts 1

Another such item is shown in Figure 9. The writer was trying to compare length and area which does not make sense in this context.

2. The length of a Square B is 4 times smaller than the area of Rectangle A. If the area of Square B is 144cm^2,

(i) Find the area of Rectangle A.
(ii) Find the perimeter of Rectangle A, if its breadth is 6cm.

Figure 9. Mathematical concepts 2

In the multiple choice item in Figure 10, the writer was not clear about the term “estimate”. The exact value 1.2 appears as one of the options which was the expected answer. This error may also be categorized as a language error in which an incorrect direction verb is used.

8 Which of the following is the best estimate of 0.8×1.5 ?

(1) 1

(2) 1.2

(3) 2

(4) 120

Figure 10. Mathematical concepts 3

In the multiple choice test item described in Figure 11, the expected answer is 250, when in actual practice the context does not allow this answer unless we can cut the given cuboid into the small cubes.

8 A cuboid measures 25 cm by 10 cm by 8 cm. How many 2-cm cubes can the cuboid hold?

(1) 10

(2) 25

(3) 250

(4) 2000 ( )

Figure 11. Mathematical concepts 4

3.3 *Errors related to diagrams as support*

The purpose of a diagram in a problem is to provide a support to the students who are going to work on the problem. In geometry, a diagram helps students to visualize the situation in which the problem is described by showing all of the relevant information pertaining to the problem concisely. If a diagram is not drawn properly, then there is a risk that it might be counterproductive making the problem more abstract for the primary students. In quite a few cases, the writers of the items produced diagrams which were lacking in several ways.

1. *Assuming sides to be perpendicular or parallel*

There were cases whereby the writers of the items assumed certain sides to be parallel or perpendicular or certain angles to be right angles. Students working on the items were expected to be able to visually identify parallel and perpendicular sides. For example, in the test item described in Figure 12, the writer assumes that the side AD of the trapezium is perpendicular to the sides AB and DC. The stem of the item did not provide relevant information about the figure.

Solve the problems below, showing your working clearly.

1. Find the area of the trapezium ABCD.

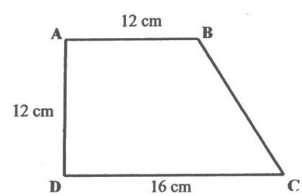


Figure 12. Assumptions about perpendicular sides

The assumption of sides being parallel and angles being right angles can be noticed in the test items described in Figures 13 and 14 below. Such assumptions can mislead students to assume these conditions from a visual examination of the figures.

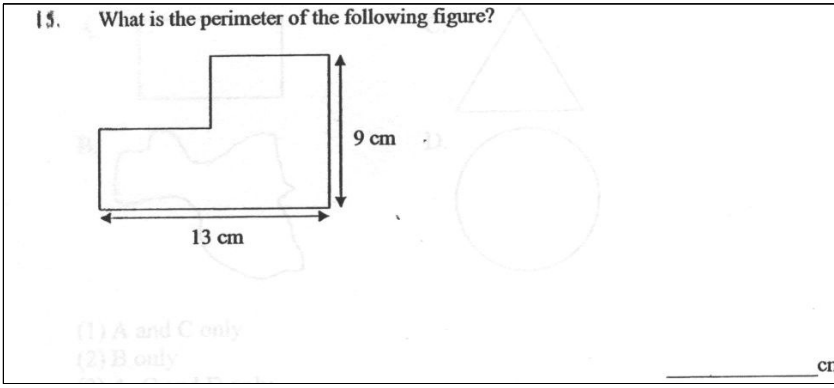


Figure 13. Assuming adjacent sides are perpendicular

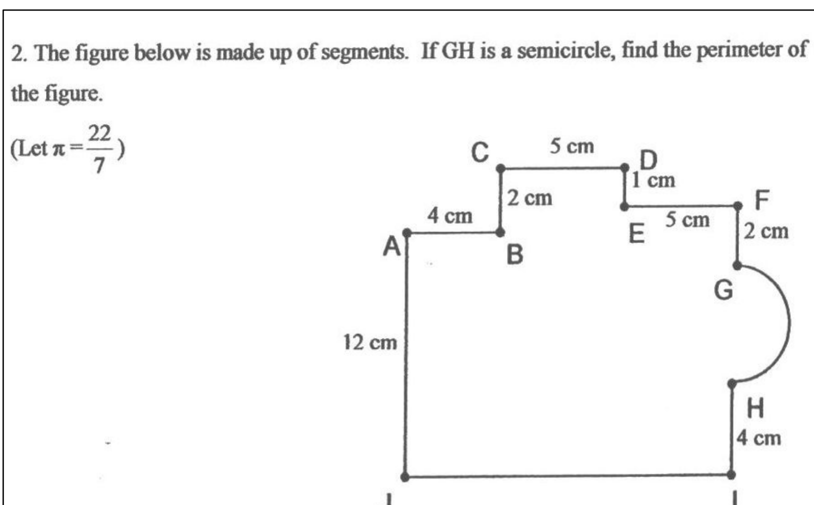
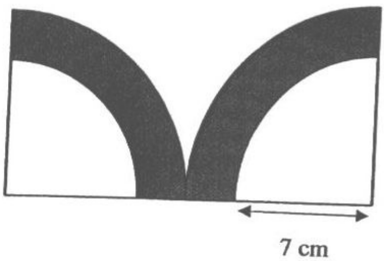


Figure 14. Assuming sides to be parallel and angles to be right angles

2. *Disproportionate diagrams*

In the test item in Figure 15 below, the writer used different dimensions in the figure to represent the same length (radius of circle). Besides using the term quadrant which is already a dificult term for primary students, the writer made the diagram quite misleading for young students.

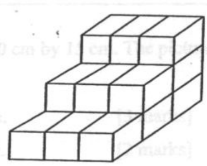


2 opposite quadrants are placed side by side as shown above. The shaded part of the quadrant is 1cm wide. Taking $\underline{\pi}$ as 3.14, find the perimeter of the shaded border.

Figure 15. Disproportionate lengths

In the test item in Figure 16, the writer showed a diagram in which the cubes look like cuboids. Although a solution can be found, a wrong idea may be conveyed to primary students about what are cubes.

10. How many unit cubes are there in this figure?



1) 6 unit cubes
2) 9 unit cubes
3) 12 unit cubes
4) 18 unit cubes

( )

Figure 16. Diagram not showing cubes

3. *Complex diagrams*

In the test item in Figure 17, the diagram was clearly very complex and misleading as well. A primary student trying to solve this problem will be completely lost as to what has to be found.

5. In the diagram below, the area of the circle, square and triangle adds up to 448 cm^2. The circle touches the square only at 4 points. Find the length of a side of the square. Take π to be $\frac{22}{7}$

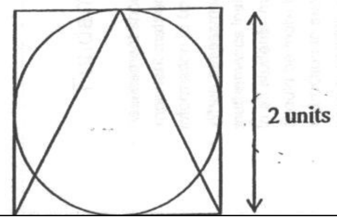


Figure 17. A complex diagram

Another item with a misleading diagram is shown in Figure 18. The writer did not make clear about the number of rectangles in the diagram.

2)
There are 3 rectangles in the figure below. Find the area labeled X when the following are given. A is 2 and half times the breadth of E. B is ¾ of the length of F. C is twice the length of D.

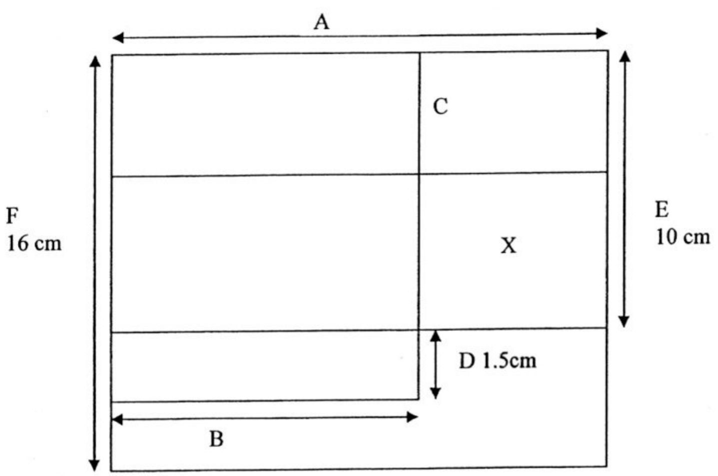


Figure 18. Misleading diagram

3.4 *Context-related errors*

The context of a problem is an important aspect that has to be carefully chosen for the problem to make sense. In the item in Figure 19, we have a very unrealistic situation whereby the 5000 m race is supposed to be run in 10 minutes and 25 seconds. It is well-known that the world record for the 5000 m race is in excess of 10 minutes and 25 seconds (currently at 12 minutes and 53.32 seconds). The context should not give wrong information. Also note the American spelling of the word *metres* used in this item.

8) Benjamin completed a 5000 m race in 10 min 25 seconds. What was his average speed in meters per second?

Ans: ____________________ m/s

Figure 19. Unrealistic context 1

In the item in Figure 20 below, the writer had ignored the fact that wooden cubes will float in water.

2. A tank, 20cm by 10cm by 5cm is filled ¼ with water. 4 wooden cubes of the same size are placed into the tank and the water level rose by 20%. When I poured in more water, the water level rose by another 20%. What is the total volume of the water and the 4 wooden cubes now?

Figure 20. Unrealistic context 2

For Figure 21, we may ask questions such as "Is the piece of cloth rectangular or not?" and "How was the cut made?"

1.	Rose bought a piece of cloth which was 12m long. She cut the cloth into 4 equal pieces. Each piece had an area of 15m^2. What was the perimeter of the cloth before it was cut?

Figure 21. Unrealistic context 3

4 Discussion

The four types of errors described above can be connected to the four categories described by Verschaffel, Greer, and De Corte (2000). For example, language-related errors can be connected to errors in the *semantic structure* as well as errors in the *format*.

All the above items came from a group of pre-service teachers. However, it is a truism that pre-service teachers are not full-fledged teachers and hence we cannot rightfully expect them to produce test items that experienced teachers will be able to produce fairly easily. Even at the end of their courses, "Pre-service teachers rarely exit their mathematics teacher preparation program as experts" (Morris, 2006, p. 471). However, teachers need to be able to recognize the qualities of an item that makes it a good item.

The issues identified above point to several causes. Carter (1984) claimed that while we wish to help students develop skills that will foster higher test results, an equal emphasis on teachers' test-making skills has not emerged. She added that: (1) there is some insecurity among pre-service teachers in writing good assessment items in a language that is still problematic for many of them, (2) the pre-service teachers tend to copy or paraphrase similar items from textbooks and their lecture notes and often make mistakes in doing so, (3) the pre-service teachers spend very little time editing or revising their test items. We may also add that not all pre-service teachers demonstrate the same level of content mastery as is required of them.

Regarding assessment, the National Council of Teachers of Mathematics (1991) has clearly mentioned that mathematics teachers should engage in ongoing analysis of teaching and learning by observing, listening to, and gathering other information about students to assess what they are learning and as well examining effects of the tasks, discourse, and learning environment on students' mathematical knowledge, skills and dispositions. So, what kinds of knowledge do teachers need for assessing their students' learning?

Shulman (1987) stated that teaching is essentially a learned profession and that teaching necessarily begins with a teacher's understanding of what is to be learned and how it is to be taught. If we assume that a teacher knows what is to be learned by the students, then the biggest issue is "how it is to be taught". This depends on several factors, of which the teacher's understanding of students' mathematical learning is extremely important. Furthermore, we can also ask: "How do we know that we have been successful in teaching?" This question has very important implications for teacher education, and we have to carefully integrate in teacher education courses knowledge about how to assess students' mathematical learning in schools. Knowledge about writing good items for mathematical tests cuts across many of the seven categories described by Shulman and most importantly mathematical content knowledge (MCK) and mathematical pedagogical content knowledge (MPCK) stand out of the lot.

If an item is poorly written, students may still feel that the item is correct and then attempt to get an answer that may make some sense within the topic area, despite the inherent flaws in the item. The assessment of students' performance based on such items is bound to be difficult to interpret. If the students apply the correct procedures and rules that they would normally apply for a well-designed test item and get the expected numerical answer, do we penalize them for not spotting the flaw in the question? If a student does not get the expected numerical answer, do we give the student the benefit of the doubt that he or she could not attempt the problem because of the inherent flaw in the problem? Thus, defective items in a test can become a very serious issue. Proper monitoring mechanism should be put in place to eliminate such items from school tests.

Pre-service teachers may improve the quality of their test items by being more systematic. A few questions that they can ask themselves (see Dindyal, 2006): What is the purpose of this item? What are the objectives to be tested? What is the mathematical topic or content area on which the item is based? Does the wording of the item correctly convey all necessary information? Are the correct direction verbs used? Is the corresponding figure associated with the item drawn correctly? Are all dimensions in the figure possible within the imposed geometrical constraints? What are the resources to be used in solving this problem? What is the expected answer? Does the answer make sense? Who will work on the items? How much time is to be spent on the solution of this problem? While this list is not exhaustive, it provides some guidance for the teacher. It is advisable that test items be pilot tested. Colleagues can also provide valuable suggestions on how to improve any test item.

5 Conclusion

Assessment has a long history in education (Van den Heuvel-Panhuizen and Becker, 2003) and will certainly be an important aspect of teaching and learning at all levels of formal education. Recent moves towards looking at *assessment of learning*, *assessment for learning,* and *assessment as learning* (see Mok, 2010) point to an increasingly important role of assessment in the curriculum. Accordingly, teachers of mathematics will continue to use tests for gauging students' learning of mathematics. Teachers cannot rely only on ready-made items from textbooks and other sources. They will have to construct original test items. One way to help teachers in assessing their students' learning would be to help them construct good assessment items. Besides well-organized professional development courses, the best way to ensure that teachers learn about tests and test construction, is to make it an essential component of their teacher preparation courses. The mathematics test items discussed in this chapter amply demonstrate that we should not take these items at face value. The items need to be carefully vetted to eliminate any shortcomings that may compromise their effectiveness. A student's response to a poorly constructed item in a test is not very informative to the teacher about what the student knows and is able to do.

References

Carter, K. (1984). Do teachers understand principles for test writing? *Journal of Teacher Education*, *35*(6), 57-60.

Dindyal, J. (2006). Defective assessment items in mathematics. *Maths Buzz*, *6*(2), 7-8.

Dindyal, J. (2009). Mathematical problems for the secondary classroom. In B. Kaur, B. H. Yeap, & M. Kapur (Eds.), *Mathematical problem solving AME 2009 Yearbook* (pp. 208-225). Singapore: World Scientific.

Kulm, G. (1979). The classification of problem-solving research variables. In G. A. Goldin & C. E. McClintock (Eds.), *Task variables in mathematical problem solving* (pp. 1-22). Washington, DC: National Institute of Education.

Lester, F. K. (1983). Trends and issues in mathematical problem-solving research. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 229-261). Orlando, FL: Academic Press.

Mason, J., & Johnston-Wilder, S. (2006). *Designing and using mathematical tasks*. St Albans: Tarquin Publications.

Mok, M. M. C. (2010). *Self-directed learning oriented assessment*. Hong Kong: Pace Publishing Ltd.

Morris, A. K. (2006). Assessing pre-service teachers' skills for analyzing teaching. *Journal of Mathematics Teacher Education*, *9*, 471-505.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.

National Research Council. (1989). *Everybody counts*. Washington, DC: National Academy Press.

Noddings, N. (1988). Preparing teachers to teach mathematical problem solving. In R. I. Charles & E. A. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 244-258). Reston, VA: National Council of Teachers of Mathematics.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22.

Van den Heuvel-Panhuizen, M., & Becker, J. (2003). Assessment design in mathematics education. In A.J .Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 689-716). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, The Netherlands: Swets & Zeitlinger Publishers.

Chapter 12

Affective Assessment in the Mathematics Classroom: A Quick Start

TAY Eng Guan QUEK Khiok Seng TOH Tin Lam

This chapter discusses the nature and rationale of affective assessment in the mathematics classroom, and introduces mathematics teachers to three techniques of affective assessment as a quick start: (1) a summated scale; (2) an interest inventory; and (3) a semantic differential. It is useful for teachers who wish to take up the challenge of assessing affects. A list for further reading is also provided for interested readers.

1 Introduction

Student assessment in the mathematics classroom often focuses on cognitive learning outcomes. However, teachers are as concerned that students develop positive attitudes towards mathematics as they are about their attainment of cognitive objectives. They lament the poor attitudes students have towards mathematics, especially those of students who do not do well in mathematics. Unfortunately, more often than not, teachers can hardly afford much time for affective development and its assessment in the classrooms. Affective learning outcomes have taken a backseat to the cognitive ones, as teachers concentrate on preparing students to do well in high-stakes examinations. It would be a challenge to carve some time out from a lesson to attend to affective learning outcomes. It would be a greater challenge if teachers are to assess for affective development in a planned or systematic manner. Instead, teachers rely on their interactions with students, the manner in which students complete the assigned work, or students' willingness to participate in learning activities, for evidence of students' attitudes towards mathematics. In fairness, even when teachers are willing, it is an unfamiliar assessment territory that they find themselves in. Compared to cognitive assessment, teachers do not talk much about techniques for assessing affect. Also, the nature of affect—consisting of abstract entities which describe behaviours characterized by feelings, emotions, interests or values—presents an assessment challenge: What is affect, precisely, and how can it be adequately measured and assessed?

In recent years, more attention has been paid to the assessment of affective learning outcomes. The Singapore Ministry of Education positions 'attitudes' as one of its five cornerstones—*Concepts*, *Skills*, *Processes*, *Metacognition* and *Attitudes*—supporting the overarching curricular goal of developing the ability of students to solve mathematics problems. Here, *attitudes* refer to the affective aspects of mathematics learning such as

- Beliefs about mathematics and its usefulness
- Interest and enjoyment in learning mathematics
- Appreciation of the beauty and power of mathematics
- Confidence in using mathematics
- Perseverance in solving a problem

(Ministry of Education, 2006)

The development of favourable attitudes towards mathematics is recognized as a crucial aim of a mathematical education. The NCTM (the National Council of Teachers of Mathematics in USA) advocates the assessment of students' mathematical disposition. In fact, two of the five general educational goals for precollege students listed in the *Curriculum Standards* (NCTM, 1989, pp. 5 – 6) are affective goals. Such assessment should seek information about students on affective objectives such as confidence in using mathematics, willingness to persevere in mathematical tasks, interest and curiosity in doing mathematics, and valuing of the application of mathematics. The Council also recommends that the evaluation of teaching should also include evidence of teachers fostering students' mathematical disposition. For example, do teachers model a disposition to do mathematics, show the value of mathematics or promote students' confidence, perseverance and curiosity in mathematics related activities?

In this chapter, we will set out some reasons for wanting to promote affective development and for assessing affect, before going on to propose some ways in which the classroom teacher can carry out affective assessment in the classroom. Along the way, we will briefly discuss the nature of affect and the challenges it presents to assessment. There are all sorts of affective targets that can be assessed; those listed in the Singapore mathematics curriculum document should suffice. Now, we know that our students know we assess what we value. By suggesting that affective assessment should be an integral part of classroom life, we hope to raise the awareness of teachers and students alike to the crucial role of affect in learning mathematics. This chapter offers the readers a small step forward to more serious dialogue among members of the mathematics education community. Readers are reminded that the sections in this chapter may be read in any order and they may proceed to the section *A few techniques of assessing affect* if they wish.

2 Some Reasons for Assessing Affective Outcomes

"My class did not do well in the final exams because it has such bad attitude towards maths." "Your son failed because he is interested in history but not maths." "This class is very enthusiastic when it comes to maths." "(Student) I don't see any use in learning factorisation of quadratic expressions; it's meaningless!" These utterances point to the affective status of students. We use the concept (or construct) of attitude, interest, or value to explain the behaviours of our students or their performance in mathematics exams. Popham (2006), who has written extensively about educational assessment and evaluation, uses the term *affect* to refer to students' attitudes, interests and values.

An immediate answer to the question of why we should assess affect is: "Students' affect should be assessed so that both teachers and students can do something about it." Affect impacts not only the immediate but also the long-term learning of mathematics. Research (e.g., Goldin, 2000; Leder, Pehkonen, and Törner, 2002; McLeod, 1992; McLeod and Adams, 1989; McLeod and McLeod, 2002) has shown the influence of affective conditions of students on students' cognitive learning. Most learning outcomes may be classified under three domains, namely, cognitive, affective and psychomotor. However, we should avoid thinking of these domains as if they are independent of each other. Positive or negative feelings, for example, can facilitate or inhibit, respectively, cognitive or psychomotor learning. This influence on learning is one reason for promoting affective learning outcomes and their assessment. A teacher who wishes to adjust the instructional approach or modify the classroom environment in order to foster the desired attitudes towards mathematics in students would in general do better with evidence from a planned assessment of attitude than from perchance observations.

The feelings and emotions one has in solving a mathematics problem, influence not only the students' immediate attainment of cognitive learning outcomes, but also their using mathematics confidently in future to solve real-world problems. Indeed, in the daily mathematics lesson, when a student is confronted with a mathematical problem (in contrast to a problem for exercise), it is a feeling of challenge or of despair that strikes the student first. It is then a matter of "fight or flight". Of course, there is a third reaction—total disinterest—which still is in the realm of affect.

Popham (2006) argues convincingly that a student's current affective status predicts that student's future behaviour and that our students' future is what we teachers are concerned about in schools. In other words, the positive feelings a student has towards solving non-routine or novel mathematics problems will predispose him or her to solve such problems in the future, which is any mathematics teacher's wish for their students' behaviour towards mathematics. In short, promoting positive affect towards mathematics is an important curricular goal for the discipline!

Teachers would want affective assessment evidence that has been systematically collected for guidance, counselling and diagnosis purposes. From experience, we know too well that a student with excellent grades in mathematics may not like mathematics enough to pursue further study. And, while we may have diagnosed the misconceptions a student has about a particular concept (think fractions) to its mathematical roots, part of our “cure” may have to address the affective aspects of learning mathematics.

Hence, affective learning outcomes are an educational goal in its own right (this is clearly stated in the MOE curriculum document) and one which should be assessed. Affective assessment of the planned and systematic variant is important for another reason. Now, being human, a teacher may think that the attitude of a certain group of students towards mathematics to be less than desirable (e.g., not persevering or interested enough). To these students, however, they are trying their utmost to do well and are frustrated by their lack of success. Simple cues from students in mathematics classrooms may suggest students’ feeling towards the subject. For example, their sullen faces which, unfortunately and by chance, may happen to be caught by a teacher, may inadvertently give the impression of a dislike for mathematics. Such a mismatch between teacher’s and students’ perception of affective status can be detrimental to both parties.

As professionals, teachers should want to ensure that our very own attitude towards our students is based on more reliable evidence than just impressions derived from unplanned observations. We want to be as certain as we can about our assessment of the students’ attitude. We do not want to unwittingly suggest to them that we think badly of their attitude towards mathematics, perhaps in a somewhat bi-directional fashion, our students’ attitude towards mathematics may be shaped by what they think their mathematics teacher’s attitude towards them as learners of mathematics is. Finally, affective assessment evidence is needed in the evaluation of the mathematics curriculum in which effectiveness is measured not only in terms of cognitive attainments but also attitudinal changes.

3 Affect: Disposition, Beliefs, Attitudes, Interests, Values and What Else?

We have to admit that we have been using the terms affect, disposition, beliefs, attitudes, interests, and values rather loosely. These concepts overlap in many ways but they are distinguishable at times. Instead of trying to define these terms using a string of words which call for further definition, we feel it will be useful at this point of our learning journey, to look into the nature of these ideas by means of examples in the next few sections. Another reason for not attempting to discuss these ideas in detail is because many people have debated and written books about them. So we will recommend some further readings (end of chapter) and leave it to the interested teachers to investigate them thoroughly.

We would call it a belief about mathematics rather than an interest if the student reports to us "I think mathematics is useful". Similarly, when the student tells us that he or she enjoys doing challenging mathematics problems, it is more an attitude towards learning mathematics rather than a belief about mathematics. If the student says that he is struggling, it would be a value rather than an indication of beliefs or interests.

Such statements are often used as prompts in a self-report or in a checklist of behaviours to reflect a particular affect. Examples of statements used in self-reports to elicit assessment evidence on the various affective targets are shown below.

BELIEFS ABOUT MATHEMATICS	*Mathematics is only understandable to geeks.*
INTEREST AND ENJOYMENT IN LEARNING MATHEMATICS	*I enjoy solving non-routine mathematics problems.*
APPRECIATION OF THE BEAUTY AND POWER OF MATHEMATICS	*Without mathematics there can be no technological advances.*
CONFIDENCE IN USING MATHEMATICS	*I am confident in using mathematics in my everyday life.*
PERSEVERANCE IN SOLVING A PROBLEM	*I give up easily when trying to solve a maths problem*

Next, we must bear in mind that attitude is always directed at an object or an idea. Quite often, when we talk about someone's attitude, we actually mean his or her attitude towards something. Hence, in the mathematics classroom, we may be interested in students' attitude towards self in learning mathematics (*I am not good at maths*) or students' attitude towards a topic (*Algebra is cool!*) and so on. There are many aspects of mathematics and its teaching and learning that can be targets for a student to have an attitude towards. We will use the term "attitudinal object" or simply "object" when we want to talk about it in general.

Finally, we cannot observe an attitude or belief directly. Beliefs, attitudes, interests, and values, being conceptual, are inferences made from observed or self-reported behaviours. We do not observe a student's attitude towards mathematics. All we do is to infer (soundly or not) from what the student does and says about or reacts to mathematics. This poses a problem: How can we be sure that our inferences from students' overt behaviours to their affective status are dependable? We shall now discuss, again briefly, the nature of affective assessment.

4 Nature of Affective Assessment

Affective assessment may involve students' opinions, preferences, attitudes, interests, and values in connection with mathematics, a topic in mathematics, learning mathematics, a particular learning activity, the mathematics teacher, or the student himself or herself as a learner of the subject. Affective assessments, compared to their cognitive counterparts, have more inherent technical and interpretative challenges. We reiterate that attitudes, interests, and values, being conceptual, are inferences made from observed or self-reported behaviours Thus, to measure and assess them, we can obtain observed evidence of affect. This can be done by observing, by using checklist, the student's behaviour related to the object (e.g., mathematics, non-routine problems), or by asking the student to report on his or her behaviour related to, or feelings towards, or views on the object. Note that we can do both but, for a busy teacher, self-reports may be more practical than direct observation of every student.

In cognitive assessments, our focus is mainly on students' *optimal* performance in tests or other assessment tasks. We assume that the students will try their best in the test within the given time. An athlete's performance in the Olympics is an example of optimal or maximal performance. For affective assessments we are more concerned about students' *typical* behaviour towards something, say, mathematics learning. Students who are positive towards self (positive self-concept) in relation to mathematics learning will have a tendency to respond favourably to learning mathematics. These behaviours are characterized by feelings, emotions, or values. Such students may, for example, show enthusiasm or engagement (covert behaviours, not observable) during the daily mathematics lesson by asking questions or volunteering to look up answers (overt behaviours, observable). Where a student typically or normally responds enthusiastically, we are inclined to say that they have a positive disposition towards mathematics. Teachers may use this information to predict how students will be predisposed to behave in the future in mathematics lessons. Likewise, students whose affect is negative (away from something, say, homework) have a tendency to respond unfavourably towards that something (e.g., dislike homework, avoid doing it). Transient or one-time feelings or emotions are of lesser concern in affective assessments, and we should guard against labelling students as having a poor attitude towards mathematics based on their atypical responses.

A technical challenge is the difficulty in linking the observed behaviours or self-reports to the concept of attitude, interest, or values. In the first place, unlike mathematics tests, there are no "right" or "wrong" answers in the assessment of affect. Depending on the manner in which the affective responses are captured, a student may react in a socially desirable way so as to project the expected image, especially where persons of authority such as their teachers are involved. Does a student's smile when solving a mathematics problem indicate enjoyment of mathematics? Well, he may be thinking what a silly question the mathematics teacher has set; or he has seen the solution of the question before. Some students may fake answers or behave untypically for some reasons, e.g., fear that their responses might be held against them. So we want students to be honest in supplying us with information about their affect. Then, it is of utmost importance that anonymity and/or confidentiality be assured in order to collect accurate affective assessment evidence.

This brings us to another aspect of affective assessment that is different from cognitive assessment for the classroom teacher. We agree with Popham (2006) that the focus of affective assessment should be on the status of the students as a class than as individuals. The requirement of anonymity and/or confidentiality, as well as respect for sensitivity of the nature of affect, necessitates a treatment of affective measures at a group or class level. Therefore we suggest, based on what we know about the nature of affect and the purpose we have for its assessment that teachers carry out affective assessment at the classroom level. We recommend that their inferences about students' affective status to be directed at students as a group rather than at a student, at least until we know how to do it more accurately.

A technical challenge in student assessment is the accuracy of the inference from the assessment evidence to the construct being assessed. For example, in a written test to assess understanding, a student's written solution may indicate memory work (hence rote-learning) rather than understanding. Hence for affective assessment, it would also be challenging to link the evidence (observed behaviours) for affective assessment to the constructs of beliefs, attitudes, interests, and values. The correlation between overt behaviours and covert attitudes or interests is far from being perfect in that we may not be able to predict accurately how a student will be disposed to respond. Being situation-specific, a student's reaction will depend on both internal (within the student, e.g., tiredness, unhappiness) and external (in the surroundings, e.g., hot day) factors at play at the time when the assessment evidence is being collected. In the prediction of behaviour from attitude, we have to constantly remind ourselves that it is a tendency or predisposition to behave in a particular way. Hence, we should avoid using once-off behaviours of students related to mathematics as their attitude towards mathematics.

5 A Few Techniques of Assessing Affect

Self-report by students and observation by teachers are two ways of obtaining affective assessment evidence. There are other ways but we recommend the self-report for classroom use, to be realistic and practical. Bear in mind that there is already little curriculum time left for affective development and affective assessment, and a teacher's work seems never done. For example, the self-report can be completed by students within a short time at the end of a lesson or taken home to complete.

5.1 *Crafting your own summative scale*

The summated scale is the most convenient technique to use in the classroom. To craft your own summated scale, follow the steps below:

1. Gather or write a number of statements relevant to the affect we have in mind. Use only statements that seem to be either definitely favourable or unfavourable to the affect object. Check your collection of statements for coherence, in the sense that the statements are all about that particular affect, so that it makes some sense to add up or "sum" the scores to each statement later. The following are statements about the affect "Attitude towards learning mathematics".

    - *I enjoy learning mathematics.*
    - *I like to work on challenging mathematics problems.*
    - *Solving mathematics problems is boring.*

2. Present the statements along with an agreement-disagreement scale to the students. Depending on a number of factors such as age of your students, and purpose of gathering the affective

evidence, the scale may consist of only two categories (Agree—Disagree) or five categories (Strongly Disagree—Disagree—Undecided—Agree—Strongly Agree). Give each category of the scale a score. For example, for a two-point scale, a score of 0 for "Disagree", and 1 for "Agree". For a 5-point scale, a score of 1 for Strongly Disagree, 2 for Disagree, 3 for Undecided, 4 for Agree, and 5 for Strongly Agree. So long as it is consistent in terms of the attitudinal direction, it makes no difference which end of the scale is given a higher score. Remark: You may have heard of the Likert scale. A Likert scale usually uses the score of 1, 2, 3, 4 and 5 for the Strongly Disagree to Strongly Agree categories.

3. Obtain the total score for a student by computing the student's score to each statement. Before doing so, make sure that the scores reflect the attitudinal direction. For example, if favourable statements are scored 5 for strongly agree, and unfavourable statements are scored 1, before computing the total score, we reverse a student's score of 1 on an unfavourable statement to 5, score 2 to 4, and leave the score of 3 unchanged, i.e., mathematically, we change a score of x to $6-x$. Remember to be consistent in the choice of attitudinal direction.

4. The total score for a student gives an indication of his or her affective status. However, recall that we should be looking at the overall affective status of the students as a group or class. Interpret the scores accordingly.

5. Remark: You may have decided to assess several aspects of affect (say, interests and values) in one go, and so put the statements about the attitudinal objects in the same self-report forms. When summing the scores, remember to add separately the scores to statements relating to a particular affect (e.g., interests) to reflect the "amount" of that particular affect; a summing up of all statements across a few affects can be misleading.

Perhaps, we may have over-simplified the process of crafting your own summated scale. We need to evaluate the suitability of the statements as indicators of the attitude we are assessing. We also have to check that the statements use a language appropriate to the target students (e.g., primary or secondary school). We should give students the instructions on how to complete and submit the self-report (here we recommend anonymity and confidentiality).

Now, you may ask: How many statements should we use? To be usable in the classroom, we recommend not having too many items. A Likert scale consists of 8 to 15 statements. However, choose a number which you think will meet your informational need. Now you can improve on the scale somewhat by carrying out what is called "item analysis". We use the idea of item discrimination which is outlined below. Generally speaking, an item or a statement must distinguish between students with positive attitude from those with negative attitude towards an object.

Form contrast groups by taking the high-scoring quarter (or half, if you must) and the lower-scoring quarter (or half) of the students who responded to the attitude scale. If there are too many students, you may reduce this to the top and bottom one third. For each statement, compare the distribution of responses of the high-group and low-group. Select statements which discriminate between the two groups. Re-write or discard the other statements. Here are some examples.

Table 1
Attitude statement: I enjoy doing non-routine mathematics problems

	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Lows	15	20	10	5	0
Highs	5	4	8	26	10

Table 2
Attitude statement: Students who don't learn mathematics miss some valuable knowledge

	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
Lows	12	20	10	4	4
Highs	10	22	8	7	6

Since about the same number in both groups agree or disagree with the second statement, that statement is NOT discriminating and should be revised or discarded. Remark: Some of you may be able to get the computer software generate these figures. Finally, if you already have your suspicions that certain students are positive in that attitude and certain others are negative, you can check if many of the students in the positive group are found in the high-scoring group, and many in the negative group are found in the low-scoring group.

Now, there can be variation in writing the attitudinal statements. Depending on the attitude or affect we are assessing, it can be interesting and relevant to use other possible response modes, such as frequency, potency, recency, or utility (Soh and Tan, 2008). Examples of the variants to the agree-disagree category are shown below:

Frequency: Asking students to report on how often.
I practise mathematics.
Almost everyday…Very seldom

Potency: Asking students to report on how fast they perform a task.
If you were given maths homework, when will you start doing it?
As soon as possible…Wait till just before handing in.

Recency: Asking students to report the last time they performed a certain task.
When did you last do extra practice (other than homework or tuition work) for maths on your own?
Last night…One month ago.

Utility: Asking students how they would use resources, e.g., spare time.
If you have to spend two hours studying, how much time will you give to maths?
None, 30 mins, 1 hour, 2 hours.

5.2 *Generating your own interest inventory*

Students are asked to indicate or check the relevant mathematics activities they are keenly interested in. They may also be asked to rank the activities in order of preference. Checking is used when the list of activities is long and ranking is suitable only when the list is short. Scoring, which depends on the purpose and nature of the measure, can be frequency counts or weighted summation.

Here is an example of asking students to rank the activities according to their preference, using 1 for most preferred and 4 for least preferred (no ties allowed).

When given a maths problem, what would you prefer to do?

	Rank
Listening to a teacher explain a problem	______
Reading a worked example of the problem	______
Trying to solve the problem on my own	______
Working on the problem in a group	______

5.3 *Crafting your own semantic differential*

For variety, you may want to consider a semantic differential. Osgood, Suci, and Tannenbaum (1957) give a list of adjectives for constructing a semantic differential. We will explain this technique by means of an example. Suppose you have just completed the topic of factorisation of quadratic expressions (secondary school mathematics) or pictorial graphs (primary school mathematics).

Choose a few pairs of diametrically opposite adjectives (or, if we may take liberties here, short phrases) that reflect the affect of interest. Ask students to mark an X in the space between each adjective pair that best indicates perception of the topic. For example

Topic (quadratic factorisation/pictographs)		
Interesting	___:___:___	Boring
Easy to learn	___:___:___	Hard to learn
Useful	___:___:___	Not useful
Like it	___:___:___	Do not like it
Not afraid of it	___:___:___	Afraid of it

You may wish to use more spaces in between the adjectives, but to be practical for classroom use, we recommend three, with X in the middle as indicating "undecided" or "no difference". Score by adding up the number of X's in each column and report counts or percentages to get an idea of how the topic went down for the students. Remember to ensure that all the adjectives are in the same direction; the adjectives above are mixed, so just adding down the columns will give the wrong answer.

Here is another example of a semantic differential with 10 items and five points used by Wong and Quek (2009).

> *Instruction to Students:* For each pair of words, circle the line that is closest to what you think or feel mathematics is.

To Me, Mathematics Is

EASY	___:___:___:___:__	DIFFICULT
BORING	___:___:___:___:___	COOL
ABOUT THINKING	___:___:___:___:___	ABOUT REMEMBERING RULES
CREATIVE	___:___:___:___:___	NOT CREATIVE
IMPORTANT	___:___:___:___:___	NOT IMPORTANT
MEANINGLESS	___:___:___:___:___	MEANINGFUL
FOR EVERYBODY	___:___:___:___:___	FOR THE GOOD ONES ONLY
NOT ENJOYABLE	___:___:___:___:___	ENJOYABLE
COMPLICATED	___:___:___:___:___	STRAIGHTFORWARD
USEFUL OUTSIDE SCHOOL	___:___:___:___:___	NOT USEFUL OUTSIDE SCHOOL

6 Conclusion

If we go by the results of international comparisons such as the TIMSS and PISA, Singapore mathematics teachers have been successful in helping students attain the cognitive goals of the curriculum. It is timely now for the mathematics teachers to attend to the relatively neglected curricular component of attitude by exploring ways of assessing affective learning in the classroom. We end here but we hope it is the beginning for mathematics teachers and educators who wish to take up the challenge of clarifying the fifth curricular cornerstone of attitude in the Singapore mathematics curriculum and assessing this goal in the classroom.

7 Further Readings

Here are some further readings. There may be other better reads and we will be happy to hear from you.

Aiken, L. R. (1996). *Rating scales and checklists: Evaluating behaviour, personality and attitude*. New York: John Wiley.

Ajzen, I. (2005). *Attitudes, personality, and behavior* (2nd ed.). Milton-Keynes, England: Open University Press.

DeVellis, R. J. (2003). *Scale development: Theory and applications*. Thousand Oaks, CA: SAGE.

Popham, W. J. (2006). *Assessing students affect.* New York, NY: Routledge. [Remark: Popham's *Mastering Assessment: A Self-Service System for Educators* is a set of fifteen practical easy-to-use booklets on educational assessment]

References

Goldin, G. A. (2000). Affective pathways and representations in mathematical problem solving. *Mathematical Thinking and Learning, 17,* 209-219.

Leder, G. C., Pehkonen, E., & Törner, G. (Eds.). (2002). *Beliefs: A hidden variable in mathematics education?* Dordrecht, The Netherlands: Kluwer Academic Publishers.

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D.A Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575-596). New York: Macmillan.

McLeod, D. B., & Adams, V. M. (Eds.). (1989). *Affect and mathematical problem solving: A new perspective*. New York: Springer-Verlag.

McLeod, D., & McLeod, S. (2002). Synthesis—Beliefs and mathematics education: Implications for learning, teaching and research. In G. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A hidden variable in mathematics education?* (pp. 115-123). Dordrecht, The Netherlands: Kluwer Academic.

Ministry of Education. (2006). *A guide to teaching and learning of O-Level mathematics 2007*. Singapore: Author.

NCTM (National Council of Teachers of Mathematics). (1989). *Principles and standards for school mathematics*. Reston, VA: Author.

Osgood, C. E., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning.* Urbana, IL: University of Illinois Press.

Popham, W. J. (2006). *Assessing students affect.* New York, NY: Routledge.

Soh, K. C., & Tan, C. (2008). *Workshop on action research: Transforming teachers to action researchers.* Hong Kong: Educational Leadership 21.

Wong, K. Y., & Quek K. S. (2009). *Enhancing Mathematics Performance (EMP) of mathematically weak pupils: An exploratory study*. Final Research Report. Singapore: Centre for Research in Pedagogy and Practice.

Chapter 13

Implementing Self-Assessment to Develop Reflective Teaching and Learning in Mathematics

Lianghuo FAN

Drawing mainly on the author's experiences in conducting research and offering courses for in-service teachers in the area of self-assessment, this chapter addresses some key issues about self-assessment, including the concepts, methods, and other related aspects about self-assessment. An exploratory study in self-assessment conducted in Singapore mathematics classrooms is also briefly presented.

1 Introduction

Over the last two decades, self-assessment, as one of relatively new assessment strategies, has received increasing attention from mathematics education researchers and practitioners in Singapore and abroad (e.g., see Brookhart, Andolina, Zuza, and Furman, 2004; Csonger, 1992; Stallings and Tascione, 1996; Wood and Frank, 2000). In Singapore, self-assessment was treated as one of the four main new assessment strategies in a major research project, the Mathematics Assessment Project (MAP), which was recently conducted under the Centre for Research in Pedagogy and Practice of the National Institute of Education (NIE) in both primary and secondary schools (see Fan, Quek, Koay, Ng, Pereira-Mendoza, and Yeo, et al., 2008). Moreover, self-assessment has also become a key topic in in-service training courses in assessment provided at NIE for school teachers over the last ten years (for a description about the training, see Fan, 2002).

In this chapter, I will draw mainly on my own experiences in conducting research and offering courses for in-service teachers in the area of self-assessment to discuss and explore some key issues about self-assessment, including the concepts, methods, and other related aspects about self-assessment. In addition, I will also introduce some research work done in this area based on the MAP project. The chapter ends with a few notes on the implementation of self-assessment.

2 What is Self-Assessment? A Conceptualisation

To better understand the concept of self-assessment, we shall first start with the concept of assessment. Assessment in mathematics is commonly defined as a process of gathering students' evidence about their knowledge of, ability to do, and disposition toward mathematics and making inferences for various purposes (see National Council of Teachers of Mathematics, 1995, p. 3), or simply, assessment is a process or act of gathering information and making inferences.

Needless to say, self-assessment literally means that one assesses him or herself (also see Van de Walle, 2004). In teachers' assessment of students in mathematics, students' self-assessment is instructed by teachers and employed by teachers to serve the purpose of teachers' assessment. More specifically, from the perspective of teachers' assessment, self-assessment is an assessment strategy where teachers gather evidence about students through their self-reviewing, self-reflecting, and self-reporting about their learning in mathematics, and hence make inferences for a variety of purposes.

For example, after a teacher has taught a chapter, say, quadratic equations, he/she can use a standard classroom test, with a suitable Table of Specifications, to check how much students have achieved about this topic. Alternatively, the teacher can also use a questionnaire survey to ask students to report to him/her whether they have understood the concepts of quadratic equations, how to solve quadratic equations using different methods, and what difficulties they still have. Both methods can serve the purposes of teachers' assessment. The former is through a traditional written test, and the latter through students' self-assessment.

The effectiveness or value of self-assessment is essentially related to the fact that, in many cases, one knows himself/herself best. In mathematics classrooms, what teachers can assess using self-assessment can be about students' cognitive or affective domain, and their learning outcomes or process (including their learning behavior).

The Singapore mathematics curriculum framework includes five inter-related aspects with the primary goal being to develop students' ability in mathematical problem solving. These five aspects are concepts, skills, processes, attitudes, and metacognition, as shown in the following well-known pentagonal structure (Figure 1).

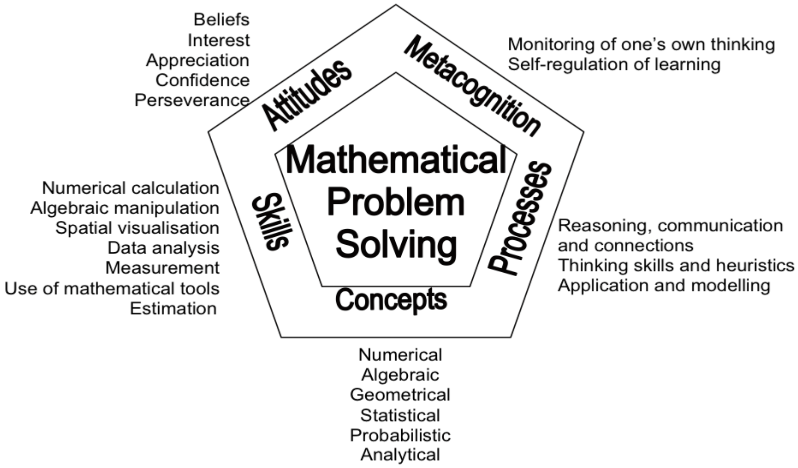


Figure 1. Framework of the school mathematics curriculum

Self-assessment can be used by teachers to gather evidence about students' learning in mathematics in all the five aspects. It is particularly effective to assess students' metacognition and attitudes, as metacognition requires students' monitoring their own thinking and self-regulating their own learning, about which the evidence cannot be easily gathered through other assessment methods (for example, written test or observation). Similarly, students' personal attitudes toward mathematics and mathematics learning, for example, whether they appreciate the usefulness of mathematics, how much they like mathematics, and how they perceive about mathematics, are best known by themselves. It seems clear that self-assessment has unique value in assessing students' metacognition and affect in the learning of mathematics. Furthermore, teachers can also reflect on or assess their own teaching, based on the information collected from students' self-assessment, and hence develop reflective teaching.

The new concept of assessment requires that teachers pay attention to not only the products (or outcomes) of learning, i.e., what students have achieved, but also the process of learning, or how students have learned. For the latter, self-assessment has also unique value, as the process of learning is essentially personalized activity and teachers need to pay attention to individual students, and for this purpose, self-assessment provides teachers with an important tool for students to report about their learning process and behavior.

Self-assessment can also help teachers better meet the challenges that come with the use of modern or reformed pedagogy, which emphasizes more student-centered learning, cooperative learning, and differentiated learning. Under these pedagogical approaches, it is easy to see that it is not effective and sometimes even impossible for teachers to gather evidence about students' learning process or behaviors by using some other assessment methods such as test and observation. Instead, self-assessment can be used as an effective tool for teachers to understand students better under these pedagogical approaches, as self-assessment allows students to reveal their own learning process and behavior.

Compared to other assessment methods, self-assessment places more responsibilities on students. It requires students to play a more active role and be their own assessors during the process of assessment. It requires students to be reflective learners by reflecting on their own mathematics knowledge, confidence, perseverance and attitudes toward mathematics and mathematics learning. In other words, not only can self-assessment help teachers, it can also help students to know better about themselves, so they can take more responsibility of their part and make necessary decision about their further learning.

It should be pointed out that the above discussion about self-assessment and its value is all from the perspective of teachers' assessment of students, namely, for teachers to gather evidence about students' learning. Hence, it is initiated and guided by teachers. However, from the perspective of learning, learners can also do self-assessment about their learning in order to make decision for their own purposes, that is, students can also be engaged in self-assessment activities in their learning process, independent of teachers' assessment. In this sense, self-assessment can be viewed as an act of students' self-reviewing and self-reflecting about their own learning in mathematics, which can lead to self-regulated learning. It appears reasonable to argue that students' engagement in self-assessment initiated by teachers will help students to develop their habit of doing self-assessment initiated by themselves.

3 How to Implement Self-Assessment in Mathematics? Methods and Examples

There are different ways for teachers to implement self-assessment in mathematics classrooms. For convenience, in this section I shall use *structured self-assessment*, *integrated self-assessment*, and *instructional self-assessment* to describe some important methods of implementing self-assessment in mathematics classrooms. I shall also use mainly the examples from the Mathematics Assessment Project (MAP) to illustrate the methods.

3.1 *Structured self-assessment*

Structured self-assessment here refers to the method that teachers conduct specific self-assessment by using pre-designed self-assessment survey forms. It is the most commonly used way for teachers to implement self-assessment.

Self-assessment survey can be used for summative purposes. In other words, it can be conducted at the end of a teaching period, for example, after completing a chapter or a topic, for teachers to know how students have learned about a chapter or topic.

In MAP, summative self-assessment is specifically implemented through a self-assessment survey form. Figure 2 below shows a standard template, called "mathematics self-assessment sheet", for teachers to use in classrooms.

Self-assessment survey can also be conducted for formative purposes, in other words, during or prior to teachers' teaching of a chapter or a topic for teachers to make informed instructional decision.

The following is a template, called "mathematics self-reflection sheet" (Figure 3), used in the MAP to gather information about students' experience in solving a particular mathematics task, and to nurture their self-reflection habit and skills.

Undoubtedly, teachers can make the necessary modifications when using these templates as self-assessment survey forms, so the assessment can be better suited to the practical situation and need in their teaching.

Mathematics Self-Assessment Sheet

Name: ___________________ Class: ________________ Date: _____________

Dear Student: We have taught (topic) last week. Please recall your experience of learning this topic and complete this self-assessment sheet, so I can know better how you have learned about this topic and how I can help you.

A. You can just circle one of the three choices: "Yes", "No", "Not sure".

1. This topic is overall easy.	Yes	No	Not sure
2. I understand this topic well.	Yes	No	Not sure
3. I had difficulty in learning this topic.	Yes	No	Not sure
4. I feel I was quite lost in learning this topic.	Yes	No	Not sure
5. I can complete homework for this topic most of the time by myself.	Yes	No	Not sure
6. I enjoy learning this topic.	Yes	No	Not sure
7. I find this topic useful.	Yes	No	Not sure
8. I am confident about this topic.	Yes	No	Not sure

B. (i) What did you find most difficult in learning this topic?

____

____

(ii) Do you still have problem about it? (Circle one) <u>Yes</u> <u>No</u>

Is there any help you wish me to give you? If so, feel free to write it down.

____

____

(iii) Did you encounter any other difficulties in learning the topic that you wish me to help? If so, please let me know.

____

____

C. What else about this topic do you wish to tell me? Please feel free to write it down (e.g., other information, or requests, or suggestions, etc.).

____

(Please use the other side of this page if necessary)

Figure 2. Mathematics self-assessment sheet (Source: MAP)

Mathematics Self-Reflection Sheet

Name: ______________________________ Date: ___________________________

1. You have tried to solve the mathematics task given below. (Note: teachers need to fill in the task, e.g., assignment, problem sum, in-class activity and exercise, etc.)

____

____

2. After doing the given task, I think (Please check the sentences that describe your work on this task)
   - € I was able to do the work.
   - € I did not understand the directions.
   - € I followed the directions but got wrong answers.
   - € I can explain how to solve this task to someone else.
   - € The task was easier than I thought it would be.
   - € The task was harder than I thought it would be.
   - € If I attempt a similar task like this next time, I will have confidence to solve it.

3. My reflection after completing the task:
   - I started the work by ____
   - I learned that ____
   - I am still confused by ____
   - I enjoyed the task because ____
   - I think the task is worthwhile because ____________________________________

4. I think the reason I made mistakes in solving this task was: (e.g., state the difficulties or problems encountered. If you did not make any mistake, please skip this item)

____

____

5. I think the most important thing (e.g., mathematics knowledge, methods, etc.) I learned from solving this task is:

____

____

6. Any other reflection? Please feel free to tell your teacher:

____

Figure 3. Mathematics self-reflection sheet (Source: MAP)

3.2 *Integrated self-assessment*

Instead of being used as a specific or independent assessment activity, integrated self-assessment refers to self-assessment that is integrated with other assessment methods. In other words, it is an integral part of an assessment package.

The recent years have seen the increasing use of many new assessment methods other than traditional written tests in mathematics. In Singapore, for example, project assessment and performance assessment have received nation-wide attention. As is well known, projects or performance tasks often require students to take an extended period of time to complete, and some are done as team work. With these approaches, it is often helpful for teachers to use self-assessment as part of the assessment package to understand how students have done with these assessment tasks.

Below is a self-assessment component (Figure 4), which is used as part of a performance assessment package, for students to do self-reflection after they have completed a performance task on mensuration at the Secondary Two level.

Self Reflection

1. What were the mathematical ideas involved in this problem?
2. Based on this activity, complete at least two of the following statements:
   - I learnt that ______________________________
   - I was surprised that I ______________________________
   - I discovered ______________________________
   - I was pleased that I______________________________
   - I am still uncertain about ______________________________
3. What have you learnt from the presentation?
4. After the presentation, how well do you assess your own performance in this activity?

Figure 4. A self-assessment component in performance assessment
(Source: Fan, 2011, p.86)

A sample of student work for the self-assessment component above is given in the appendix of this chapter. Readers who are interested to know more about the performance task and its assessment rubrics can refer to Fan (2011).

Similarly, self-assessment can also be integrated with other assessment methods, for example, the whole or part of a journal writing assessment task can be designed for students to write about their self-assessment. It can also form part of the portfolio assessment about their learning over a period of time. Nevertheless, a further discussion about this aspect is beyond the intention of this chapter.

3.3 *Instructional self-assessment*

Instructional self-assessment is an ongoing self-assessment that is embedded in teachers' classroom instruction. It can be treated as part of the teachers' daily instructional activities, especially classroom discourse with the whole class or individual students. It is usually not structured and pre-designed. Instead, it is often impromptu and instantaneous in the context of instructional practice.

In the MAP project, the following list of prompts (see Figure 5) is used for teachers to engage students in self-assessment with different instructional scenarios.

4 What Does Research Tell Us? An Exploratory Study in Self-Assessment

In the earlier sections, we have discussed the concepts, methods and other related issues of self-assessment in mathematics classrooms. In this section, we shall briefly introduce an exploratory study that was recently completed on the use of self-assessment in mathematics classrooms in Singapore[1]. It should be pointed out that, as a recent

[1] Readers who are interested to know more details about the study or about the MAP project can refer to Fan, Teo, & Pereira-Mendoza (2009) or the MAP technical report (Fan, Quek, Koay, Ng, Pereira-Mendoza, & Yeo, et al., 2008).

review revealed, there have been very few classroom-based studies available in the area of self-assessment (see Fan, Quek, and Ng, et al., 2006).

Prompts for Instructional Self-assessment for Teaching Problem Solving

Prompt 1	Where did you encounter difficulties? Why? (Scenario 1: after students did not know how to start or proceed in solving a problem)
Prompt 2	Where did you go wrong? Why? (Scenario 2: after students realized that he/she got a wrong solution or answer)
Prompt 3	Is the mistake a careless mistake? If not, why did you make the mistake? (Scenario 3: after students realized that he/she made a mistake)
Prompt 4	Are you sure your answer/solution is correct? Did you check? (Scenario 4: after students solved a problem or finished a task)
Prompt 5	Have you solved this kind of problems before? Does the problem look familiar to you? (Scenario 5: when students encounter difficulty in solving a problem, which appears to be essentially not new to him/her in his/her learning)
Prompt 6	What have you learned from solving this problem? (Scenario 6: after students has gone through an important or difficult problem)
Prompt 7	If you are given another problem like this, will you have confidence to solve it? (Scenario 7: after students solved a problem in a correct way)

Prompts for Instructional Self-assessment for Other Teaching Scenarios

Prompt 8	What did you feel most difficult in learning this chapter (or topic, or lesson, or task, etc.)? Is it still difficult to you? (Scenario 8: after students finished learning a chapter, a topic, a lesson, a concept, etc.)
Prompt 9	Do you have any questions or difficulties to ask? (Scenario 9: When teacher prepares to close his/her teaching for chapter, a topic, a lesson, a section, a task, a problem, etc. and moves to next phase)
Prompt 10	How do you feel about your learning of mathematics recently? Do you think you can improve your learning? How? (Scenario 10: when teacher realized that students might have problems recently in learning mathematics)

Figure 5. A list of prompts for instructional self-assessment (Source: MAP)

The study was part of a larger research project, the Mathematics Assessment Project (MAP), which involved a classroom-based intervention in 16 primary and secondary schools for about three school semesters. As mentioned earlier, four relatively new—assessment strategies were studied in the MAP project. These four strategies included project assessment, performance assessment, self-assessment, and communication-based assessment. One of the main reasons for the researchers to include self-assessment in the study was to let students be more responsible for their learning, so they could be better motivated and be more reflective in their learning.

The study had two main research questions. First, what were the influences of using self-assessment strategies on students' learning of mathematics in their cognitive domain? Second, what were the influences of using self-assessment strategies on students' learning of mathematics in their affective domain? By focusing on these two questions, the researchers hoped to better understand how the use of self-assessment strategy can be effectively integrated into mathematics classrooms in Singapore.

The study was carried out in eight mathematics classrooms, including four Primary Three and four Secondary One classrooms in two primary and two secondary schools (two classes in each participating school). For convenience and to be more specific, below we shall focus on one secondary school, which is a high-performing school in terms of the average students' achievements in GCE O-level examinations.

The intervention of self-assessment was mainly implemented through the following tasks or activities: 1. Student self-evaluation, which requires students to evaluate their own learning at the end of a period of teaching, usually once a week or once a topic; 2. Student self-reflection, conducted as and when it was needed; and 3. Self-assessment prompts, which were designed to be used during teachers' daily classroom discourse when appropriate and helpful under different scenarios. Figures 2, 3 and 5 given above show some intervention tools for classroom use in these three aspects.

The research data, including both quantitative and qualitative, were collected in a variety of ways, including pre- and post- questionnaire surveys, school-based examinations, pre- and post- self-assessment tests, self-assessment intervention tasks, classroom observation, and interviews with participating students and teachers.

While the quantitative data showed that there appeared to be neutral or statistically non-significant influence of the self-assessment on students' achievement on school-based examinations, the qualitative data, particularly the interviews with both teachers and students, showed that self-assessment can not only help teachers improve their teaching by responding to students' reflection and feedback, but also promote students' self-awareness and meta-cognition in their learning of mathematics by making them think harder and deeper about their own learning, and hence help them become better reflective learners.

The study found that the participating teachers were very confident about the use of self-assessment in their classroom teaching. They believe that self-assessment is a good strategy that would benefit teaching and learning of mathematics and it can be integrated into their mathematics classrooms. In particular, teachers felt that self-reflection can be done regularly as a routine activity. In fact, one of the participating teachers started doing so with non-participating classes after gaining the necessary experience during the study. Similarly, all of the students interviewed also supported the idea of using the self-assessment strategy in their classrooms.

Regarding implementation issues, the results suggest that, while self-assessment is quite feasible to implement with both teachers and students, it should still be used wisely, including making the requirement of self-assessment tasks clear to the students, giving students adequate instructions and help, and using it regularly but not too frequently.

From the exploratory study, the researchers concluded that effectively implementing self-assessment strategies in the mathematics classrooms can not only help teachers understand better students' learning and their own teaching, but also provide students with meaningful opportunities to reflect on their own learning, and hence improve teachers' teaching and student' learning.

The overall findings from all the four schools implementing self-assessment were quite consistent, particularly in students' and teachers' views and reactions about self-assessment, as shown in the qualitative data collected. Based on the findings of the study, the researchers argued that student self-assessment can and should be done as an integral or routine activity in the teaching and learning of mathematics.

5 Concluding Remarks

To end this chapter, I would like to point out that, in order to effectively implement self-assessment in mathematics teaching and learning, teachers must create a positive and encouraging learning environment, so students are not afraid of telling the truth about their learning in mathematics. Obviously, this kind of learning environment is vital for teachers to understand students' learning difficulties, frustrations, and needs for help through the use of self-assessment.

In addition, teachers should also realize that the evidence gathered from students through self-assessment is only one indicator about their learning in mathematics. Due to different reasons, it is possible that sometimes students might not be willing, or able, to tell the truth (for example, students might not know what they do not know, or they might over- or under- estimate their learning difficulties). Therefore, it is important for teachers not only to design or use effective self-assessment tools, but also to help students' develop reflective skills. It is also helpful, and sometimes even necessary, that teachers use other assessment methods to gather evidences about students' learning for triangulation purpose and, ultimately, for a more valid and reliable assessment.

References

Brookhart, S. M., Andolina, M., Zuza, M, & Furman, R. (2004). Minute Math: An action research study of student self-assessment. *Educational Studies in Mathematics, 57,* 213-227.

Csonger, J. E. (1992). Sharing teaching ideas: Mirror, mirror on the wall. . . Teaching self-assessment to students. *Mathematics Teacher, 85,* 636-640.

Fan, L. (2002). In-service training in alternative assessment with Singapore mathematics teachers. *The Mathematics Educator*, *6*, 77-94.

Fan, L., Quek, K. S., Ng., J. D., et al. (2006). *New assessment strategies in mathematics: An annotated bibliography of alternative assessment in mathematics*. Singapore: Centre for Research in Pedagogy and Practice (CRPP), National Institute of Education.

Fan, L., Teo, S. W., & Pereira-Mendoza, L. (2009). *Student self-assessment for better learning in mathematics: An exploratory study in Singapore classrooms*. Paper presented at the 5th International Self Conference, Al Ain, UAE.

Fan, L., Quek, K. S., Koay, P. L., Ng, J., Pereira-Mendoza, L., Yeo, S. M., et al. (2008). *Integrating new assessment strategies into mathematics classrooms: An exploratory study in Singapore primary and secondary schools*. Singapore: Centre for Research in Pedagogy and Practice, National Institute of Education. Retrieved January 28, 2011, from http://www.crpp.nie.edu.sg/~pubs/CRP24_03FLH_FinalResRpt.pdf

Fan, L. (Ed.) (2011). *Performance assessment in mathematics: Concepts, methods, and examples from research and practice in Singapore classrooms*. Singapore: Pearson.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

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

Van de Walle, J. A. (2004). *Elementary and middle school mathematics: Teaching developmentally* (5th ed.). Boston, MA: Pearson.

Wood, D. K., & Frank, A. R. (2000). Using memory-enhancing strategies to learn multiplication facts. *Teaching Exceptional Children*, *32*, 78-82.

Appendix

A sample of student work on a self-assessment component in performance assessment

Self Reflection

1. **What were the mathematical ideas involved in this problem?**
   - The use of formulas to find areas and volume of different figures
   - The calculation of a solid with a "hollow" middle
   - measurements of volume the lengths of different sides of the chocolate had to be made

2. **Based on this activity, complete at least two of the following statements:**
   - I learnt that assumptions are inevitable in calculating volume and surface area of the chocolate.
   - I was surprised that I a chocolate's shape and structure affects the consumer.
   - I discovered that the same amount of chocolate would have a smaller surface area when made into a cube rather than a sphere
   - I was pleased that I got to eat the chocolates and I love chocolates
   - I am still uncertain about the actual surface area and volume of the chocolates because of the assumptions we made.

3. **What have you learnt from the presentation?**
   - There will be some assumptions made in calculating the surface area and volume but some methods are more accurate than others. There may also be different methods to calculate the surface area of a chocolate. The fewer the number of assumptions made, the more accurate the answer.

4. **After the presentation, how well do you assess your own performance in this activity?**
   - I made a few more assumptions than the groups who presented, making my calculations easier but not as accurate. However, the calculated figures are still approximately the same and so the comparison of taste, density and shape could still be carried out ~~as~~ rather accurately.

5. What I like most about this Performance Task?

   I was able to eat the chocolates after making the measurements and I noticed many more details about what was in the chocolate bar than I did before

Contributing Authors

CHAN Chun Ming Eric is a lecturer at the Mathematics and Mathematics Education Academic Group in the National Institute of Education, Nanyang Technological University. He lectures on primary mathematics education in the pre-service and in-service programmes. His research interests include children's mathematical problem solving, mathematical modelling, problem-based learning, and learning difficulties.

David John CLARKE is a Professor at the University of Melbourne and Director of the International Centre for Classroom Research (ICCR). Over the last fifteen years, his research activity has centred on capturing the complexity of classroom practice through a program of international video-based classroom research. The ICCR provides the focus for collaborative activities among researchers from China, the Czech Republic, Germany, Hong Kong, Israel, Japan, Korea, New Zealand, Norway, the Philippines, Portugal, Singapore, South Africa, Sweden, the UK, and the USA. Other significant research has addressed teacher professional learning, metacognition, problem-based learning, and assessment (particularly the use of open-ended tasks for assessment and instruction in mathematics). Current research activities involve multi-theoretic research designs, cross-cultural analyses, discourse in and about classrooms internationally, curricular alignment, and the challenge of research synthesis in education. Professor Clarke has written books on assessment and on classroom research and has published his research work in around 140 book chapters, journal articles, and conference proceedings papers.

Jaguthsing DINDYAL holds a PhD in mathematics education and is currently an Assistant Professor at the National Institute of Education in Singapore. He teaches mathematics education courses at both the primary and secondary levels to pre-service and in-service school teachers. His interests include geometry and proofs, algebraic thinking, international studies, and the mathematics curriculum.

Lianghuo FAN holds a Personal Chair in Education at the School of Education, University of Southampton, UK, where he also serves as Head of the Mathematics and Science Education Research Centre. He obtained his MSc at East China Normal University, Shanghai and his PhD at the University of Chicago, USA. From 1998 to 2010, he was on the faculty in the Mathematics and Mathematics Academic Group of the National Institute of Education of Nanyang Technological University, Singapore during which he had also served as the principal investigator of the Mathematics Assessment Project (MAP) and later as principal investigator of the Singapore Mathematics Assessment and Pedagogy Project (SMAPP), both under the Centre for Research in Pedagogy and Practice, NIE. Professor Fan's research interest includes mathematics instruction and assessment, teacher professional development, curriculum development, and algorithm of polynomial algebra in mathematics. He has published widely in these areas, including an internationally acclaimed book, *How Chinese Learn Mathematics: Perspectives from Insiders* (co-edited with Wong, Cai, and Li). Professor Fan was chief editor of *The Mathematics Educator*, and is currently editor-in-chief of two series of mathematics textbooks published in Singapore and China.

Marja van den HEUVEL-PANHUIZEN is a Professor of Mathematics Education at the Freudenthal Institute for Science and Mathematics Education, Utrecht University, the Netherlands. At this university she obtained her PhD in Mathematics Education with a study on assessment in mathematics education. The heart of her work is in the further development of the didactics of mathematics as a scientific discipline.

Her research interests are largely focused on the area of primary school mathematics. Her special interest is assessment. Other key topics of her research activities are gender differences in mathematics education and the use of didactical models to support the learning of mathematics. She has worked on a number of national and international research projects, including comparative and evaluative studies, design research, and classroom experiments. Presently, she is involved in research on the use of picture books to support kindergartners' learning of mathematical concepts, the use of ICT for teaching primary school students early algebra, the use of dynamic tools in e-assessment to identify weak students' learning potential, the use of games in mathematics education, context-based assessment and the analysis of mathematics textbooks. In the winter term of 2004-2005, she was a visiting professor at Dortmund University. From November 2005 to the end of 2009, she was a visiting professor at the Institut für Qualitätsentwicklung im Bildungswesen (IQB) at Humboldt University Berlin.

JIN Haiyue is currently a PhD student with the Mathematics and Mathematics Education Academic Group, National Institute of Education, Nanyang Technological University, Singapore. She obtained her Master degree in mathematics education from Nanjing Normal University, China, in 2007. Her research interests include mathematics assessment and the teaching and learning of mathematics at the secondary level. Her current PhD study focuses on students' conceptual understanding in mathematics and the use of concept map as an assessment technique.

Berinderjeet KAUR, PhD, is a Professor of Mathematics Education and Head of the Centre for International Comparative Studies (CICS) at the National Institute of Education in Singapore. Her primary research interests are in the area of classroom pedagogy of mathematics teachers and comparative studies in mathematics education. She has been involved in numerous international studies of Mathematics Education and is the Mathematics Consultant to TIMSS 2011. She is the principal investigator (Singapore) of the Learner's Perspective Study (LPS) helmed by Professor David Clarke of the University of Melbourne. As the President of the Association of Mathematics Educators (AME) from

2004-2010, she has also been actively involved in the professional development of mathematics teachers in Singapore and is the founding chairperson of the Mathematics Teachers Conferences that started in 2005. She is also the founding editor of the AME Yearbook series that started in 2009. On Singapore's 41st National Day in 2006, she was awarded the Public Administration Medal by the President of Singapore.

Angeliki KOLOVOU is a PhD student at the Freudenthal Institute for Science and Mathematics Education (FIsme) of Utrecht University. She obtained her Master's degree in Mathematics Education from the Kapodistrian University of Athens, Greece. Her research is about mathematical problem solving by (upper) primary school students. In particular, she investigates students' solving processes in non-routine problems and ways to support these processes.

LEONG Yew Hoong is an Assistant Professor with the Mathematics and Mathematics Education Academic Group, National Institute of Education, Nanyang Technological University, Singapore. He has a continuing interest in classroom research and mathematics teacher education. He enjoys working with mathematics teachers to test out instructional ideas within the realistic constraints of classroom teaching.

Magdalena Mo Ching MOK is Chair Professor of Assessment and Evaluation of the Department of Psychological Studies, and Co-Director of Assessment Research Centre at The Hong Kong Institute of Education.

Marjolijn PELTENBURG is a PhD student at the Freudenthal Institute for Science and Mathematics Education, Utrecht University in the Netherlands. Her PhD research focuses on the development and investigation of assessment instruments for disclosing special education students' potential in mathematics. She will finish her PhD thesis in 2012. Peltenburg has worked at the Freudenthal Institute as a curriculum developer in special education from 2004-2008. She has a background in educational science.

QUEK Khiok Seng is a Senior Lecturer with the Psychological Studies Academic Group, National Institute of Education, Nanyang Technological University, Singapore. His areas of study focus on assessment and teacher education. He is an active collaborator on projects that investigate the integration, influence, and use of new assessment strategies in mathematics classrooms, including the recent Mathematics Assessment Project funded by the Centre for Research in Pedagogy and Practice, NIE. He is also the co-principal investigator of the project Mathematical Problem Solving for Everyone (M-ProSE).

Yoshinori SHIMIZU, PhD, is a Professor of Mathematics Education at University of Tsukuba, Japan. His primary research interests include international comparative study on mathematics classrooms and assessment of students learning in mathematics. He is the Japanese team leader of the Learner's Perspective Study (LPS), a sixteen countries comparative study on mathematics classrooms. He was a member of Mathematics Expert Group (MEG) for OECD/PISA and has been a member of the Committee for National Assessment of Students Academic Ability in Japan.

TAY Eng Guan is an Associate Professor with the Mathematics and Mathematics Education Academic Group, National Institute of Education, Nanyang Technological University, Singapore. He obtained his PhD in Mathematics (Graph Theory) from the National University of Singapore. Dr Tay continues to do research in mathematics as well as in mathematics education, particularly in the area of problem solving. He has papers published in international scientific journals in both areas and has co-written a number of undergraduate level textbooks on Counting and Graph Theory. Dr Tay has taught in two junior colleges in Singapore and had a stint in the Ministry of Education before he joined the National Institute of Education.

Denisse R. THOMPSON is Professor of Mathematics Education at the University of South Florida in the United States. She received her PhD in Education from the University of Chicago in 1992, conducting an evaluation of the 12th grade course (*Precalculus and Discrete*

Mathematics) developed by the University of Chicago School Mathematics Project (UCSMP); she has been involved with UCSMP for over 25 years, as an author, editor, and most recently as Director of Evaluation for the Third Edition materials. Dr. Thompson taught at the middle school, high school, and community college levels before joining the faculty in her current position in 1990. Her major scholarly interests are curriculum development and research, literacy in mathematics, and the integration of culture into the teaching of mathematics. She has authored or co-authored 17 books, over 25 book chapters, and over 50 journal articles and presented at over 100 conferences, including the International Congress on Mathematics Education, annual and regional meetings of the National Council of Teachers of Mathematics, and the East Asia Regional Conference on Mathematics Education. In 2004, she was named the Kenneth Kidd Educator of the Year by the Florida Council of Teachers of Mathematics, and in 2010 was named the Mathematics Teacher Educator of the Year by the Florida Association of Mathematics Teacher Educators.

TOH Tin Lam is an Associate Professor with the Mathematics and Mathematics Education Academic Group, National Institute of Education, Nanyang Technological University, Singapore. He obtained his PhD in Mathematics (Henstock-stochastic integral) from the National University of Singapore. Dr Toh continues to do research in mathematics as well as in mathematics education. He has papers published in international scientific journals in both areas. Dr Toh has taught in junior college in Singapore and was head of the mathematics department at the junior college level before he joined the National Institute of Education.

WONG Khoon Yoong has worked as a mathematics educator in Malaysia, Australia, Brunei Darussalam, and Singapore. His research and writings cover mathematics teacher education, comparative education, ICT in mathematics teaching, and learning strategies. He is currently the principal investigator for Singapore in the IEA TEDS-M study (Teacher Education and Development Study in Mathematics), an APEC study on high school mathematics and science teacher education, and the

Singapore Mathematics Assessment and Pedagogy Project (SMAPP) under the National Institute of Education, Nanyang Technological University, Singapore.

YEO Kai Kow Joseph is currently a faculty member at the Mathematics and Mathematics Education Academic Group in the National Institute of Education, Singapore. He has been a mathematics teacher educator for more than 10 years, teaching mathematics education courses at both the primary and secondary levels to pre-service and in-service school teachers as well as heads of mathematics departments. Before joining the National Institute of Education in 2000, he held the post of Vice Principal and Head of Mathematics Department in secondary schools. His research interests include mathematical problem solving in the primary and secondary levels, mathematics pedagogical content knowledge of teachers, mathematics teaching in primary schools, and mathematics anxiety.