

Using Performance Assessment in Secondary School Mathematics: An Empirical Study in a Singapore Classroom

Lianghuo Fan

Nanyang Technological University, Singapore

Yan Zhu

University of Hong Kong, Hong Kong

This article reports an exploratory study on using performance assessment in mathematics instruction in a high-performing secondary school in Singapore. An intact mathematics class participated in the study, and received chapter-based performance tasks as intervention during regular mathematics lessons for about one and a half school years. The performance tasks used included authentic and/or open-ended tasks. The students' academic achievements and attitudes in mathematics were compared with a comparison class that did not receive the intervention. Both quantitative and qualitative data were collected, mainly through questionnaire surveys, performance task tests, conventional school exams, and interviews with students and teachers. The results suggest that the students receiving the intervention performed significantly better than their counterparts in solving conventional exam problems, and in general they also showed more positive changes in attitudes towards mathematics and mathematics learning. The students from the experimental class also expressed positive views about the benefits of using performance tasks in promoting their ability in higher order thinking, though no statistically significant difference was detected between the two classes of students in solving unconventional tasks before and after intervention. Overall, the results appear to support teachers' using contextualised problems in real life situations and open-ended investigations in students' learning of mathematics.

Key words: performance assessment, alternative assessment, mathematics teaching and learning, authentic questions, open-ended questions, Singapore mathematics education.

Introduction

Over the last decades, the importance of assessment in education, and particularly in mathematics education, has received much attention from educational policy makers, mathematics education researchers, and mathematics teachers, among others, in many countries (e.g., see Fan, 2006; Morgan, 2000; National Council of Teachers of

Mathematics [NCTM], 1995; Niss, 1993; Ruthven, 1994; Singapore Ministry of Education, 2004a, 2004b).

In particular, relatively new assessment (or the so-called alternative assessment) concepts and strategies have been widely advocated by educational reformers and increasingly used by practitioners in classroom practices (e.g., see; Brookhart, Andolina, Zuza, & Furman, 2004; Clarke, 1997; Hargreaves, Earl, & Schmidt, 2002; Kulm, 1994). Along with new development, assessment is more viewed as an integrated part of the teaching and learning process. As Pegg and Panizzon (2007/2008) noted, “The emphasis on embedding assessment into the teaching and learning process is identifiable globally”.

A relatively large research project focusing on integrating new assessment strategies into mathematics teaching and learning was recently conducted in Singapore schools. The new assessment strategies under investigation in the project include those using communication tasks, performance tasks, project tasks, and student self-assessment tasks. Each new assessment strategy was implemented in two primary and two secondary schools, in total 16 schools, for about one and half school years. The effects of using the new assessment strategies on students’ learning of mathematics were studied in both affective and cognitive domains.

As part of the large project, the study presented herein focuses on the effects of using performance tasks, including mainly authentic and open-ended ones, in mathematics instruction on students’ learning of mathematics in one participating secondary school. The study was intended to address the following three research questions:

- (a) What are the effects of using performance tasks on students’ mathematics performance as measured in solving unconventional problems?
- (b) What are the effects of using performance tasks on students’ mathematics performance as measured in solving conventional problems?
- (c) What are the effects of using performance tasks on students’ attitudes toward mathematics and mathematics learning?

It is hoped that the study can provide research-based evidence on the potential influences of using performance tasks on students’ mathematics learning so as to help school teachers better align assessment practice with the desired educational goals and hence improve the quality of teaching and learning.

Conceptual Framework and Perspectives

Performance assessment has been widely believed to have more pedagogical value and it can reflect students’ achievement more accurately than traditional multiple-choice tests (e.g., see Kane, Khattri, Reeve, & Adamson, 1997). According to NCTM, assessment is “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” (1995, p. 3). Following this definition, ‘performance assessment’, or sometimes called ‘performance-based assessment’, is an assessment strategy by which the evidence about students’ learning is gathered through students’ work on performance tasks. Nevertheless, there is no consensus on its definition, or more specifically, what performance tasks are. According to Buechler (1992), the emergence of

the performance assessment movement was due to the fairly widespread dissatisfaction with high-stake multiple-choice tests. Gripps (1994) claimed that, in the United States, performance assessment was often regarded as any type of evaluation which was not multiple-choice or standardised testing. However, such a definition is rather broad and it covers almost all types of alternative assessment (e.g., project work).

In the Third International Mathematics and Science Study (TIMSS), performance assessment was included as one important component in international comparison, and was referred to as integrated and practical tasks targeting students' content and procedural knowledge as well as their ability in using knowledge for reasoning and problem solving (Harmon et al., 1997). The Wisconsin Education Association Council (1996), at the root of the meaning of the word 'performance', set their definition as the one requiring students to demonstrate skills and competencies by performing or producing something. The central idea in Stenmark's (1991) definition about performance assessment is to assess what students actually know and can do. It is clear that while researchers tend to differentiate performance assessment from traditional assessment, they have different concerns and focuses. In other words, a variety of aspects have been connected with the term performance assessment.

To be more applicable to the Singapore school education context, performance tasks used in this study were mathematical tasks with the following two distinguishing characteristics: (a) authentic in context, and (b) open-ended in approaches and answers. Naturally, when students solved these tasks, they were required to demonstrate how they performed, in other words, their thinking or working process.

The authenticity of a problem, according to the NCTM, is the degree to which tasks are faithful, comprehensive, and complex, which can be found in important, real-life performances of adults that are non-routine yet meaningful and engaging for students (NCTM, 1995). It is believed that tasks with this feature could engage students in applying knowledge and skills they have learned in the classroom to real-world challenges, and help them appreciate the usefulness of mathematics.

The open-endedness of a problem includes two aspects: (a) multiple venues of access or ways of solutions, and (b) multiple acceptable answers to the problem. It is believed that solving open-ended problems is more challenging than close-ended ones that students usually encounter in their school work, and normally requires higher-order thinking. In fact, these two aspects are to a large degree lacking in traditional assessment tasks, which consequently often have received criticism over the last decades (e.g., see Howe & Jone, 1998; Wu, 1994).

As a result, all the performance tasks used in this study are contextualised, to a different degree, in real-world scenarios; they can be approached in various ways and ended with different answers (not just in different representation forms). Below is a sample performance task, authentic in the Singapore social context. More examples of performance tasks designed and used in the study can be found in Fan (2008).

Use the information listed in *Distribution by Type of Dwelling* as shown below to (1) construct a pie chart of the distribution of dwelling types in 2000; (2) predict and construct a pie chart of the distribution in 2010 and defend your answer.

Distribution by Type of Dwelling

| % | 1980 | 1990 | 2000 |
|------------------------------|------|------|------|
| HDB Flats ¹ | 68.5 | 84.6 | 88.0 |
| Condominiums & Private Flats | 2.3 | 4.1 | 6.0 |
| Private Houses | 8.5 | 7.0 | 5.1 |
| Others ² | 20.7 | 4.3 | 0.9 |

¹Government-built apartments. ²Includes shophouses, attap/zinc-roofed houses, other public flats and others.

This task introduces students to the knowledge about various housing types in Singapore and the changes in the distribution over the last three decades. The task requires students to predict the possible distribution in ten years later based on the information given. To have a reasonable prediction, students need to use their knowledge in constructing pie charts and apply their daily life experience to figure out a possible changing trend in the housing types and relevant factors that may have influences on the changes. Among the four housing types, the changes for *condominiums & private flats* and *private houses* appear stable over the last thirty years, while the changes for the other two types appear more rapid over the same period, which may in particular lead some students to think of the possibility of the disappearance of the housing type *others* in the next ten-year time frame. Given that there are no universal and standard answers to the question, three performance criteria were given for the teacher to assess students' answers: (a) the trend of change must be consistent with the previous two time intervals, (b) all the percentages in the pie chart should sum up to 100%, and (c) the arguments for the predictions must be reasonable and practical.

Like in many other countries, developing students' ability in solving authentic and open-ended problems in the teaching and learning of mathematics has received increasing attention in Singapore. As a matter of fact, the Singapore mathematics syllabus emphasises the importance of students' applying mathematics in solving real-life problems and being engaged in open-ended investigations in mathematics instruction (Ministry of Education, 2002). However, an analysis of two widely-used Singapore secondary mathematics textbooks revealed that fewer than 2% of textbook tasks were authentic and about 2% were open-ended (Fan & Zhu, 2000, 2007). Moreover, according to Schoenfeld (1992), the beliefs that mathematics learning has little or no relation to the real world and that any mathematics task has one and only one answer are commonly held among students. In this connection, the study has both theoretical and practical significance.

Research Methods Below we shall provide information about the participants, including the school, students and their mathematics teachers, the instruments for data collection, as well as the procedures of data collection and data analysis in this study.

Participants

As mentioned earlier, this study focused on one participating secondary school, which is identified as a high-performing school, as it was randomly selected from the 50

best performing secondary schools according to year 1999 to year 2002 GCE “O” Level Examination results released by the Singapore Ministry of Education.

Thirty-eight Secondary One (Grade 7) students from one intact class of high ability in the high-performing school were selected in this study to receive chapter-based interventions on performance tasks during regular mathematics lessons for about three school semesters starting from early 2004. A parallel intact class of 40 students was chosen as a comparison group. Table 1 provides the profiles of the students and their mathematics teachers from the two classes. No significant difference was found between the two classes in terms of students’ Primary School Leaving Examination (PSLE) overall scores ($t [76] = 0.81, p = 0.42$) and mathematics grades ($U [38, 40] = 747.00, p = 0.84$).

As we can see from Table 1, the two classes were taught by two different teachers with basically equivalent professional background since the beginning of the study in 2004. However, it should be noted that due to some unforeseen reasons, starting from January 2005 the teacher teaching the experimental class had to take over the comparison class as well. Given the change, the teacher was advised not to use the intervention tasks in the comparison class so as to keep the teaching practices unchanged in both the classes in terms of interventions.

Table 1
A Profile of Students and Mathematics Teachers in the Experimental and Comparison Classes

| | Experimental Class | Comparison Class ¹ |
|-------------------------------|--------------------------|-------------------------------|
| No. of Students | | |
| Boys | 24 | 25 |
| Girls | 14 | 15 |
| Mathematics teachers | | |
| Gender | Male | Male |
| Length of teaching experience | 3 months | 9 months |
| Qualification | M.Eng, PGDE ² | MSc |

Note. ¹ In year 2005, the teacher teaching the experimental class also took over the comparison class; ² PGDE stands for Postgraduate Diploma in Education.

The teacher teaching the experimental class received training and guidance on how to use performance tasks in teaching before and during the intervention from the researchers. Intervention tasks were carried out in the experimental class throughout the intervention period of about one and half years, while the comparison class was taught as usual during this period of time.

Instruments and Data Collection

Four main instruments were designed for data collection in this study: questionnaires, performance task tests, intervention task worksheets, and interviews with teacher and students.

Questionnaires

Two questionnaires, one for the pre-intervention survey and the other for the post-intervention survey, were designed to find out students' attitude toward mathematics and mathematics learning as well as their experience with performance tasks (see Appendix). Both questionnaires consist of two parts, Part I and Part II. For a comparison purpose, Part I, containing 22 items, is the same for both questionnaires. Those items, focusing on students' perceptions about mathematics and their learning of mathematics, can be categorised into four subgroups based on the areas they measured: general view towards mathematics and mathematics learning, anxiety level in mathematics learning, perceptions of own performance in mathematics, and beliefs about the usefulness of mathematics. A nine-point Likert-type scale ranging from "disagree totally" to "agree totally" is employed for each item.

Part II in the pre-intervention questionnaire was intended to measure students' experience with various alternative assessment tasks. Out of the 6 items in total, 3 items are particularly about their experience in solving performance tasks in their mathematics learning before intervention. Each item is in a six-point Likert-type scale on frequency. Part II in the post-intervention questionnaire focused on students' feeling about using performance tasks in the learning of mathematics. There are totally 16 items, each using the nine-point Likert-type scale.

A pilot test of the pre-intervention questionnaire was conducted in January 2004, involving 56 Secondary One students from two other schools. The initial version of the questionnaire was modified based on the results of the pilot tests. The final questionnaire has a reliability of 0.85.

The pre-intervention questionnaire survey was conducted in February 2004 for both the experimental and comparison classes with a response rate being 100% and the post-intervention questionnaire survey was in May 2005 with a response rate being 81.6%.

Performance task tests

Similar to the questionnaires, two sets of parallel performance task tests, a pre-test and a post-test, were designed. The pre-test was intended for the researchers to have a better understanding about students' entry levels in mathematics problem solving, while the post-test was conducted for the researchers to measure the possible changes of students' ability in problem solving after three school semesters with or without being exposed to performance tasks in mathematics learning for experimental and comparison classes, respectively. Both tests contain three open-ended tasks, with one being also authentic.

A pilot study of the pre-test was conducted in February 2004 with 35 Secondary One students from one school. Based on the students' feedback, necessary modifications on test tasks were made. The modified tasks were again piloted by a group of 36 Secondary

One student from another school in March 2004. As a result, while about 60% of the students felt that the tasks were challenging to them, all the students had no difficulty in understanding the tasks. Some minor modifications were further made in finalising the pre-test items.

By the way, students' performance in solving conventional tasks was measured using normal school exam scores. With the participating teachers' assistance, we were able to collect all the 78 students' PSLE overall scores and mathematics grades (Exam A), year 2004 school mid-year mathematics exam scores (Exam B), year 2004 school final-year mathematics exam scores (Exam C), as well as year 2005 school first mathematics common test scores (Exam D). In addition, the pre-test, which focused on students' ability in solving performance tasks, was conducted in March 2004 with a response rate being 97.4% and the post-test, parallel to the pre-test as mentioned earlier, was in May 2005 with a response rate being 82.4%.

Intervention task worksheets

For the purpose of integrating performance tasks into classroom teaching and learning, the design of the intervention tasks strictly followed the stipulated school scheme of work. Moreover, all the intervention tasks meet both the criteria as described earlier: authentic as well as open-ended. For each chapter covered in the scheme of work, one to two performance task worksheets were first crafted by the researchers and then finalised jointly by the researchers and the participating teacher to better match the students' background and the teacher's teaching plan.

In total, the teacher of the experimental class managed to carry out a total of 12 interventions during the three school semesters. Students' work on each intervention task was collected by the classroom teacher and then handed to the researchers for evaluation. After grading, a copy of students' work with researchers' comments was returned back to individual students for their information and possible revision.

The researchers observed most interventions to monitor how the performance tasks were carried out in the classroom. During those classroom observations, the interventions were recorded with field notes, or audio/video taping. The observations were also useful for the researchers to improve the design of future performance tasks.

Interviews with teacher and students

The interviews with teacher and students were conducted in late May 2005, after all the surveys and tests mentioned above were completed. The purpose of the interview was mainly for getting information about the participants' experience and understanding regarding use of performance tasks as well as their opinions or suggestions on the use of the new strategy in teaching and learning. While the interview questions for both teacher and students are similar, understandably those for teacher are more from a perspective of teaching and those for students are more from a perspective of learning.

A total of three interview sessions were carried out. The mathematics teacher from the experimental class received an individual interview, while six students (two high performing, two average performing, and two low performing) were recommended by the teacher to attend the student interviews, in which these students were grouped in three for

each session. The two student sessions lasted about 30 minutes each and the one with the mathematics teacher was about 60 minutes. All the interviews were recorded by audio taping and filed notes.

Data Process and Analysis

The data from the two questionnaires were analysed using quantitative methods. Descriptive statistics (e.g., frequency and percentage) was applied to describe students' overall perceptions about mathematics and mathematics learning. Mann-Whitney U tests were used to examine the possible differences between the two classes of students in each survey for the researchers to detect the impact of using performance tasks on the experimental students' attitudes.

Students' work in the two performance task tests was graded based on task-specific rubrics by two independent researchers. The inter-rater reliability was calculated by the Intraclass Correlation Coefficient (ICC) on absolute agreement. As a result, the reliability on three performance criteria (i.e., Approaches, Solutions, and Representation) over the three tasks for the two tests ranged from 0.98 to 1.00, with an average being 0.99. Similar to the analysis for the questionnaire data, the rubric-based grades from the performance task tests were analysed by descriptive statistics to investigate students' overall performance at class levels before and after the intervention period. Mann-Whitney U tests were employed to identify possible differences between the experimental and comparison classes in each test. Wilcoxon Signed-Ranks tests were used to detect the change of students' grades from the pre- to post-tests. Moreover, possible differences on the changes between the two classes were examined by Mann-Whitney U tests to identify the potential relationship to the intervention program.

Students' PSLE overall scores and mathematics grades in the experimental and comparison classes were compared by t-tests and Mann-Whitney U tests respectively to measure the equivalence of students' academic background in the two classes. The following three normal school exam scores were analysed by 2×2 ANOVA with time (Exam B vs. Exam D; Exam B vs. Exam C; Exam C vs. Exam D) as a within-subject factor and treatment (experimental vs. comparison) as a between-subjects factor to investigate potential effects of using performance tasks on students from the experimental class.

The interview data collected in an audio format were transcribed. Using qualitative methods, the data allow researchers to analyse the teacher and students' views about the new type of assessment strategy, which is not easy to be gained by questionnaire surveys or achievement tests. Moreover, the evidence from the interview is helpful for the researchers to triangulate what has been revealed in the above quantitative data so as to strengthen the findings of the study.

Limitations of the Study

Like many other intervention-based studies in educational research, understandably there were also difficulties and hence limitations in this study, given that it was conducted in authentic classroom settings.

First, to investigate the effects of using performance tasks on students' learning of

mathematics, this study involved one experimental class and one parallel comparison class. Ideally, unlike the experimental class, the comparison class should not be exposed to performance tasks during the intervention period of about three school semesters. However, as mentioned earlier, for the last school semester, the mathematics teacher of the experimental class had to take over the comparison class due to unforeseen reasons, which was beyond the control of the researchers and the teacher. Although having worked with the researchers, the teacher clearly understood and agreed that he should not introduce the ideas of performance assessment (let alone intervention tasks) to his teaching in the comparison class, the experience of the teacher working with the experimental class could still, though likely unintentionally, influence his teaching in the comparison class one way or another, which can, to an extent, affect the results of the study.

Second, while this study introduced performance tasks to the teaching and learning of mathematics in the experimental class, at the school level those students were still assessed based on the traditional assessment practice for their official school performance grading and reporting. In other words, the new assessment strategy was only introduced at the classroom level, but not at the school level, which could affect students' motivation in their working on the new assessment tasks, and therefore have negative influences on the results of the study (see more discussions in the next section).

Third, according to the research design, the experimental students should be exposed to the performance tasks in a systematic and scheduled way. However, due to some unexpected school activities, it was often very difficult for the teacher to do so in delivering the tasks to the students. In particular, in the first semester, the class only managed to carry out one intervention task, nevertheless with continuous efforts of the researchers, the teacher, and the school administrators, the situation was significantly improved in the second and third semesters.

Given those limitations, we wish to remind the readers that the conclusions of study should be taken with some caution. More generally, we wish to emphasise that the study should be viewed as an exploratory one, which was also our intention when we designed the study.

Results and Discussions

The main findings of the study were reported below, based on the three research questions mentioned earlier.

Effects of using performance tasks on students' mathematics performance in working on unconventional problems

As said earlier, the pre- and post- performance task tests were targeted at students' mathematics performance in solving unconventional problems. Similar to the intervention tasks, all the tasks in the pre- and post- tests are open-ended in approaches and answers, and one task in each test is contextualised in a real-life scenario. Moreover, the tasks in the post-test were designed to be parallel (equivalent) to those in the pre-test, hence the researchers can better measure the difference in students' performance and how

their experience with the new assessment tasks during the intervention affect their performance in the post-test.

In terms of the overall scores, the Wilcoxon Signed Ranks tests showed that the students from both the experimental and comparison classes made significant improvement from the pre- to post-test (Experimental: $Z = 3.60$, $p < 0.001$, $r = 0.62^1$; Comparison: $Z = 3.88$, $p < 0.001$, $r = 0.75$). No significant difference between the two classes was detected in either test.

As all the tasks are open-ended in nature, it appears more meaningful to further examine students' performance in terms of the effective strategies they employed, the number of answers they obtained, and the solutions they represented. In fact, these aspects are the three performance rubrics designed for evaluating students' performance in the tests. A brief description of the three performance rubrics is listed in Table 2.

Table 2
A Brief Description of General Rubrics by Approaches, Solutions, and Representation

| | Level 0 | Level 1 |
|--|---|--|
| Approaches (decision/strategy about Approaching tasks) | No attempt or No evidence of a strategy | Strategy is ineffective and could not lead to any correct answer |
| Solutions (no. of answers obtained) | No correct answer obtained | Only ne orrect answer obtained |
| Representation (documentation of problem solving procedures) | No attempt or Working is irrelevant | Working is not clear and hard to read |
| Level 2 | Level 3 | Level 4 |
| Strategy could lead to correct answer but not systematic (e.g., guess and check) | Strategy shows partial systematic pattern | Strategy is ffective That would lead to acomplete set of Answers |
| More than one correct answer obtained | At least 50% of the full answers obtained | A complete set of answers obtained |
| Working is not organised so that the approach is not observable | Working is organised and approach is partially observable | Working is well organised and approach is fully observable |

It is believed that the analysis on the sub-domains can provide more in-depth information on how students approach and solve such challenging unconventional mathematics problems, especially those from the experimental class.

Regarding the approaches employed by the students, the data revealed that in most cases, the students were able to use more systematic/effective methods in the post- than pre-test. That is, more students received a mean score over 2 on this performance scale in the post- than pre-test (Experimental: 79.4% vs. 42.1%; Comparison: 82.8% vs. 23.7%). The improvements in both the classes reached a significant level (Experimental: $Z = 3.93$,

¹ Effect size r is calculated when significant difference is detected. According to Cohen (1992, 1988), an r value over .5 is considered to be 'large', around .1 to be 'weak', and around .3 to be 'medium'.

$p < 0.001$, $r = 0.67$; Comparison: $Z = 4.12$, $p < 0.001$, $r = 0.79$), but no significant difference was found between the classes in terms of the improvements.

As indicated earlier, all the tasks in the tests contain more than one correct answer, as listed below:

| | Pre-Test | Post-Test |
|--------|----------|-----------|
| Task 1 | 7 | 10 |
| Task 2 | 56 | 25 |
| Task 3 | 2 | 2 |

Since the last task in each test only had two answers, a task-specific rubric on solutions was set for the two tasks, shown as follows:

| Task 3 | Level 0 | Level 1 | Level 2 | Level 3 | Level 4 |
|---|---------------------------------------|---|--|--|---|
| Solutions (no. of answers obtained) | • No correct answer obtained | • Only partial correct answer obtained, i.e., getting correct central number(s) Or • Answers obtained just by switching surrounding numbers without changing the central numbers | • One complete answer with different central number obtained | • Two correct central numbers with one complete answer obtained | • Two complete answers with different central numbers obtained |

The analysis revealed that compared to the pre-test, the percentages of students who stopped at obtaining one correct answer (i.e., average score ≤ 1.33) were much smaller in the post-test for both the classes. In fact, the two classes of students made significant improvement in getting multiple correct answers from the pre- to post-test (Experimental: $Z = 3.39$, $p < 0.001$, $r = 0.58$; Comparison: $Z = 3.79$, $p < 0.001$, $r = 0.73$). However, a between-class comparison did not display any significant difference regarding the improvements as well as students' performance on the particular performance rubrics in either test.

It is believed in the study that representation is also an important skill in problem solving. Therefore, although it is not a focus of the intervention program, how students represent their solutions in the performance task tests was examined. The results revealed that the students generally did not have significant changes in their representation from the pre- to post-test. Moreover, consistent with the results on the other two performance rubrics, the two classes of students did not have significantly different performance on the aspect of representation in either the test.

Overall, it was found that both the experimental and comparison students made significant improvement from the pre- to post-test, especially in the aspects of using effective strategies and getting multiple correct answers. At first, the results appeared somehow disappointing to us as it did not favour the experimental class. However, the interviews with the teacher and students revealed that students' working on performance tasks overall had a positive influence on their learning of mathematics. For instance, one

student felt that the open-endedness of performance tasks gave him a different view of maths, say, “math is not just doing homework questions”. One student stated that doing the performance tasks “allow us think differently”. A few students commented that the new type of tasks help them “to think out of the box” instead of “sticking to one way”. As to the teacher, although he expressed some concerns about some weak students, he maintained that such tasks are generally beneficial for his students’ learning in mathematics, especially for “strong class [students] ... who have the potential in them by their own”.

A further discussion among the researchers and the teacher, to some degree, provides some explanations to the above seemingly inconsistent findings. According to the teacher, the students from the experimental class knew well about the research and they were clear that all their grades on performance tasks would not be counted into their school records. It might have affected their motivation in the post-test so that it is possible that these students did not treat the test as seriously as their peers from the comparison class who were just given the tests without further information. On the other hand, the result might also imply that developing students’ ability to a higher level in solving challenging performance tasks could take a longer time than we have expected. In this regard, further study is needed before we can make a definite conclusion, which is beyond the scope of the current study.

Effects of using mathematics performance tasks on students’ mathematics performance in solving conventional problems

Concerning students’ performance in solving conventional mathematics problems, we used students’ PSLE mathematics grades and end-of-semester assessment scores throughout the intervention period. As reported earlier, there was no significant difference in the students’ PSLE overall scores as well as mathematics grades (Exam A) between the experimental and comparison students, which provided an indicator of equivalence between the two classes. In Exam B, the equivalence still remained ($t [75] = 0.02, p = 0.99$). In fact, till the year 2004 mid-year school examination, the experimental class only managed to implement one intervention. Therefore, no great change for the experimental class was expected. More interventions were carried out later on, as shown in Figure 1.

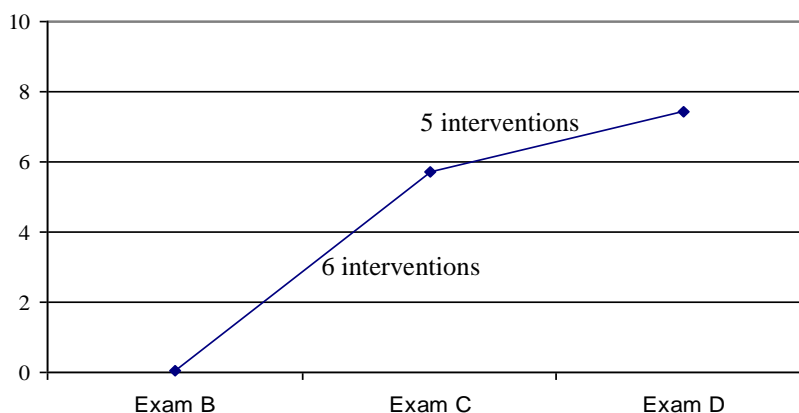


Figure 1. Difference of the average scores between the experimental and comparison class on school exams.

From the figure, we can easily find that the differences between the two classes increase rapidly in the last two school exams. In particular, the two differences reached a considerably significant level (Exam C: $t [74] = 1.94, p = 0.06, r = 0.22$; Exam D: $t [74] = 1.96, p = 0.05, r = 0.22$). A repeated measurement analysis of variance between Exam B and Exam D revealed that there was a significant interaction between time and treatment effects ($F [1, 74] = 8.39, p < 0.005, r = 0.32$) and the effect size is about medium, which is in favour of the experimental class. A further analysis revealed the significant interaction actually occurred between the period from Exam B and Exam C ($F [1, 73] = 6.68, p < 0.005, r = 0.29$), and in the next period (from Exam C to Exam D), the experimental class held the superiority. It appears clear that the students from the experimental class had an advantage over the comparison students when doing their conventional school exam tasks, which suggests that students' exposure to performance tasks is beneficial for them to solve conventional tasks, an implicit hypothesis we had at the beginning of the study. It is also interesting to investigate how long the positive influence would maintain, which is, however, beyond the scope of the study.

Effects of using performance tasks on students' attitudes toward mathematics and mathematics learning

The data about students' attitudes towards mathematics and mathematics learning are collected through the pre- and post-questionnaire surveys. As said earlier, Part I in both questionnaires was the same, with focus on their general perceptions about mathematics and mathematics learning, and the items in Part I can be categorised into four subgroups.

The first subgroup, consisting of six items, was about students' general views about mathematics and their learning of mathematics. The data revealed that the two classes of students overall provided positive responses to these items in both the surveys. However, it was also found that the students in both the classes became more negative in the post- than in the pre-survey in terms of average rating².

In general, there were no significant differences between the two classes of students in terms of their general views toward mathematics and mathematics learning, though the differences were in favour of the experimental class in both the surveys. Nevertheless, the difference between the two classes became smaller in the post- as compared to the pre-survey. In particular, while the experimental students appeared significantly more willing to spend time in studying mathematics than the comparison students in the pre-survey (Item 16: $U [38, 39] = 552.50, p < 0.05, r = 0.25$), the responses between the two classes had no significant difference in the post-survey (Item 16: $U [33, 29] = 363.50, p = 0.10$). We think a possible reason for this change is that the students from the experimental class

² This result was not surprising, as available research has found that students at lower grade levels often have more positive views about mathematics because of a variety of reasons, for example, mathematics becomes more challenging to students at higher grade levels (e.g., see Macnab & Payne, 2003; Wong, Lam, Wong, Leung, & Mok, 2001).

had more opportunities to work on performance tasks, therefore they might have developed a perception that they had made enough effort and spent enough time in studying mathematics. Another possible reason is that these students might have experienced more frustrations because of solving challenging performance tasks, which were reflected in the interview. For example, one student stated that “[we] have to gather a lot of information and use it and sometimes ... we don’t [know] where to put what in the question” and another said that “you do a lot of work, you still cannot solve, you are very irritating”.

The second subgroup, also consisting of six items, was about students’ anxiety level in learning mathematics. While the students from the experimental class gave overall positive responses to all the relevant items in the two surveys, those from the comparison class only provided positive response to all the items in the pre-survey, but not in the post-survey. In particular, the comparison students expressed in the post-survey that they were somehow under terrible strain in mathematics lessons (Item 2) and not confident when it came to mathematics (Item 20).

The results also showed that the experimental students were consistently less anxious about their mathematics learning than the comparison students, though both the classes became more anxious from the pre- to post-survey. As can be found from Table 3, there are statistically significant differences in students’ responses to a number of items between the two classes in the post-survey. Specifically, the experimental students were significantly less stressed (Item 2), less afraid of (Item 6), less nervous (Item 17), and more confident about mathematics (Item 20) than their counterparts and the effect sizes on the four items ranged from 0.33 to 0.42 with an average being 0.38. In comparison, no significant difference was found in the pre-survey.

Table 3
**Comparison Between the Experimental and Comparison Classes
on Anxiety Level Items**

| | Pre-Survey (<i>U</i>) | Post-Survey (<i>U</i>) |
|---------|-------------------------|--------------------------|
| Item 2 | 654.50 | 293.50** |
| Item 6 | 675.00 | 297.00** |
| Item 10 | 677.50 | 357.00 |
| Item 14 | 600.00 | 403.50 |
| Item 17 | 570.00 | 322.00* |
| Item 20 | 645.00 | 278.00** |

Note. * $p < 0.05$, ** $p < 0.01$; The *U* values in the tables are obtained by Mann-Whitney U-test, which examines the differences in the ranked positions of ratings between the experimental and comparison classes.

The fact that the experimental students have been exposed to performance tasks appears to be one reason for such a result. As the performance tasks are generally more challenging than normal school mathematics tasks, the experimental students then had more opportunities to be engaged in higher-order thinking via working on those tasks. Therefore, these students became less anxious about mathematics for both challenging tasks as well as normal school tasks.

The third subgroup of items was about students' perceptions about their own performance in mathematics. It also comprises six items. The data revealed that the students were happy with their own performance, however unlike experimental students, the comparison students indicated in general that they did not like solving challenging mathematics problems in the post-survey (Item 21). A comparison of the two classes of students' responses to this item showed that the experimental students were significantly more willing to attempt challenging mathematics tasks than the comparison students in the post-survey ($U [33, 29] = 333.00, p < 0.05, r = 0.30$) but no significant difference was found in the pre-survey ($U [38, 40] = 671.00, p = 0.37$). It appears that the experience with performance tasks did help the experimental students to develop a positive disposition towards working on challenging tasks.

However, we also found that while the experimental students had significantly stronger beliefs that they could do well in mathematics (Item 15) than the comparison

students in the pre-survey ($U [38, 39] = 546.50, p < 0.05, r = 0.26$), such a difference did not show again in the post-survey ($U [33, 29] = 398.50, p = 0.25$). This result is not surprising to us, because the performance tasks the experimental students had worked on during the intervention period were generally not easily solvable, and this fact would likely lead the experimental students to better appreciate the challenging nature of mathematical tasks and hence give a more modest answer, compared to their counterparts in the comparison class whose responses may only refer to the normal school mathematics tasks they had encountered.

The fourth subgroup comprised four items designed to examine students' beliefs about the usefulness of mathematics, which is related to one particular feature of the performance tasks used in this study: authenticity in the task context. The results showed that the students from both the classes provided overall positive responses on all the relevant items in the two surveys and no significant between-class differences were detected in either survey. However, compared to the pre-survey, the students' responses in the post-survey became more negative for both the classes. Such a change is quite understandable, as in general mathematics became more abstract and appeared further away from students' daily life when students moved to higher grades. In particular, we noticed that for the experimental students, the negative change on the item about the meaningfulness of studying mathematics reached significant level (Item 12: $U [38, 33] = 436.00, p < 0.05, r = 0.30$). This result may be related to the fact that while the experimental students were given many opportunities to work on performance tasks which involved real-life application of mathematics knowledge, the skills they learned from the new assessment strategy, however, were seldom assessed in their formal school examinations. This inconsistent practice may lead students to believe that working on performance tasks was somehow a waste of time. The interviews also revealed a number of such thoughts. For example, one student commented that "some [tasks] are not relevant to our normal maths" and the teacher told us that "none of these are tested in the academic tests, ... so ... the kids also sometimes question whether ... they need to spend so much time on these ... the first question they ask ... 'is this counted in CA1?'" These

results reminded us of the importance of aligning teaching and assessment, especially for those students who are test-oriented.

Part II of each of the questionnaires was targeted on students' experience with new assessment strategies, including authentic tasks as well as open-ended tasks. Particularly for the experimental students, items were added to measure their perceptions about performance tasks in the post-survey.

The data from the pre-survey showed that both the experimental and comparison classes had overall similar experience in doing the tasks with the aforementioned features. Basically, they worked on the tasks with the relevant features either on a monthly basis or a weekly basis. The comparison students' responses to the same items in the post-survey were not significantly different from those in the pre-survey, which indicates that the teaching practice in the comparison class remained unchanged in terms of the use of performance tasks and it is consistent with the research design.

Regarding the experimental students' new experience with performance tasks, the results from the post-survey revealed that they generally accepted well the specific features of the performance tasks, including multiple approaches of the tasks (Item 26) and the authenticity in task contexts (Item 30 & Item 31). Moreover, the students believed that doing performance tasks helped them to be more creative (Item 27) and systematic (Item 32). However, it seems that, to a degree, the students were still uncomfortable with the open-endedness in final answers. In the interviews, some students also told us about their confusions. For example, one student told us, "because you have found one of the answers, and then if we check with other people for the answers, we thought that either one of us was wrong". It is understandable that in their previous school experience, students were often merely required to provide one and only one correct answer to a task and they had already been used to such practice in solving mathematical tasks and felt comfortable with it.

The experimental students generally felt that doing performance tasks was very challenging. More than 60% of the students claimed that they had to think harder in doing the tasks (Item 28), 58% believed that it was time-consuming (Item 35), about 21% felt lost in doing the tasks (Item 29), and 36% needed hints when working on those tasks (Item 33).

In terms of the usefulness of doing performance tasks, the results showed that the majority of the students did not have negative views toward such experience; in particular, they believed that doing those performance tasks could help them in learning mathematics (Item 25) and made them learn mathematics better (Item 37). However, about one third of students felt that working on performance tasks was a waste of time (Item 38) and only slightly more than one tenth of the students were willing to take more performance tasks in their future learning (Item 36). Such a result could be related to the fact that, as mentioned earlier, performance tasks were not included in the formal school exams. Therefore, some of the experimental students were unable to "see" the immediate benefit of doing the tasks for at least it does not seem to help them to get higher marks in the conventional school tests. Consequently, these students did not fully see the usefulness of doing performance tasks in their mathematics learning and hence became unwilling to have more in future study.

Conclusions and Implications

This study aimed to investigate the effects of integrating performance assessment into regular school mathematics teaching and learning. In the study, performance tasks were defined as those contextualised in real-world settings, approachable using various methods, and with multiple acceptable answers. Thirty-eight Secondary One students in an intact class from one randomly selected high-performing Singapore school received a three-school-semester intervention, with the other forty students as an intact comparison class. By using questionnaire surveys, performance task tests, students' normal school exam scores as well as interviews with teacher and students, the researchers investigated the possible impact of using performance tasks on the experimental students' mathematics learning in both academic and affective aspects.

Regarding students' academic achievement, the study looked into students' performance in both the conventional assessment (school exams) and unconventional assessment (i.e., performance task tests). The results in the normal school exams showed that the changes in students' performance across three continuous school semester tests were significantly preferable in the experimental classes. Moreover, the favourite changes occurred after the intervention program had been implemented about one school year, where the experimental class completed 7 interventions, and maintained till the interventions ended. Although it is hard to attribute the positive result solely to students' experience with performance tasks, it appears clear that the students from the experimental classes did benefit from being exposed to performance tasks.

In the unconventional tests, the students from both the experimental and comparison classes performed significantly better in the post- than pre-tests, not only in terms of their overall scores but also in specific performance domains, including using effective strategies and obtaining multiple answers. However, no significant difference was detected between the two classes regarding their progresses. As suggested earlier, one possible reason for such a result is that the students from the experimental class well knew that their performance in the test would not be counted into their school records, which might have limited their performance in the test. The result might also imply that developing students' ability to a higher level in solving challenging performance tasks could take a longer time than we have expected. In this regard, further evidence is needed before we can make a definite conclusion. Nevertheless, it is clear that no negative effect of using the new assessment strategy was found on students' performance.

Consistent with many other researchers' findings (e.g., Macnab & Payne, 2003; Wong, Lam, Wong, Leung, & Mok, 2001), the students in this study in general become more negative toward mathematics and mathematics learning in the post- from the pre-survey. On the other hand, the study also revealed that the changes were generally in favour of the students from the experimental class, especially in the anxiety level about mathematics. However, it appears not expected that the changes on the view about the usefulness of mathematics were preferable to the students from the comparison class, while contextualisation in real life was one important characteristic of performance tasks, and the experimental students expressed their appreciation for the specific features of performance tasks in Part II of the post-survey as well as interviews. One possible reason

for the seemingly contradicting results is that the students may find that what they experienced in the performance tasks seldom appeared in their regular school mathematics learning. To them, working on normal school mathematics tasks was more important, as it would really be tested. Correspondingly, the experimental students may have an even stronger feeling that the mathematics they encountered in the regular school learning was farther from their daily life.

The results from the study suggested that teachers and students were capable of handling performance tasks. Although the effects of using the new strategy in some cases were not obvious, the study did observe positive effects on students' academic achievement and their anxiety level about mathematics learning. Moreover, it is clear that no negative impact was observed. In short, the overall results appear to support teachers' using contextualised problems in real life situations and open-ended investigations in students' learning of mathematics.

Finally, we would like to point out that this study was an initial step for us to explore the possible effects of using performance tasks on both teachers' teaching and students' learning of mathematics. More research on the impact of using such new strategies on teaching and learning in various aspects is needed. Moreover, given the complexity of the practice of teaching and learning, there is a long way for us to go to fully understand how the new assessment strategy can be effectively used to improve the quality of teaching and learning, particularly with different students (e.g., with different abilities).

Acknowledgement

The study reported in this article was funded by the Ministry of Education of Singapore through the Centre for Research in Pedagogy and Practice (CRPP), National Institute of Education, Nanyang Technological University, Grant No. CRP24/03FLH. All the views expressed are solely those of the authors and do not reflect the views of the CRPP. The authors also wish to thank Mr. He Ming for his technical assistance in the final stage of completion of this article.

References

- Brookhart, S., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics*, 57, 213-227.
- Buechler, M. (1992). Performance assessment. *Policy Bulletin*, No. PB-B13.
- Clarke, D. (1997). *Constructive assessment in mathematics: Practical steps for classroom teachers*. Berkeley, CA: Key Curriculum Press.
- Fan, L. (2006). Making alternative assessment an integral part of instructional practice. In P. Y. Lee (Ed.), *Teaching secondary school mathematics: A resource book* (pp.343-354). Singapore: McGraw Hill.
- Fan, L., & Zhu, Y. (2000). Problem solving in Singaporean secondary mathematics textbooks. *The Mathematics Educator*, 5(1/2), 117-141.
- Fan, L., & Zhu, Y. (2007). Representation of problem-solving procedures: A comparative look at China, Singapore, and US mathematics textbooks. *Educational Studies in Mathematics*. 66(1), 61-75.

- Fan, L. (Ed.) (2008). *Performance assessment in mathematics: Concepts, methods, and examples from research and practice in Singapore classrooms*. A resource booklet presented for the 2008 seminar organized by the Curriculum Planning and Development Division of the Ministry of Education for the Heads of Departments of Mathematics in secondary schools, Singapore.
- Gripps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *Educational Researcher*, 39(1), 69-95.
- Harmon, M., Smith, T. A., Martin, M. O., Kelly, D. L., Beaton, A. E., Mullis, I. V. S., et al. (1997). *Performance assessment in IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.
- Howe, A. C., & Jones, L. (1998). *Engaging children in science* (2nd ed.). Upper Saddle River, NJ: Merrill.
- Kane, M. B., Khattri, N., Reeve, A. L., & Adamson, R. J. (1997). *Assessment of student performance*. Washington, DC: Studies of Education Reform, Office of Education Research and Improvement, U.S. Department of Education.
- Kulm, G. (1994). *Mathematics assessment: What works in the classroom*. San Francisco: Jossey-Bass.
- Macnab, D. S., & Payne, F. (2003). Beliefs, attitudes and practices in mathematics teaching: Perceptions of Scottish primary school student teachers. *Journal of Education for Teaching*, 29(1), 55-68.
- Morgan, C. (2000). Better assessment in mathematics education? A social perspective. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 225-242). Westport, CT: Ablex.
- National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.
- Niss, M. (1993). (Ed.). *Cases of assessment in mathematics education: An ICMI study*. Dordrecht, The Netherlands: Kluwer.
- Pegg, J., & Panizzon, D. (2007-2008). Addressing changing assessment agendas: Impact of professional development on secondary mathematics teachers in NSW. *Mathematics Teacher Education and Development*, 9, 66-80.
- Ruthven, K. (1994). Better judgment: Rethinking assessment in mathematics education. *Educational Studies in Mathematics*, 27, 433-450.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook on mathematics teaching and learning* (pp. 334-370). New York: Macmillan.
- Singapore Ministry of Education. (2002). *Mathematics syllabus (lower secondary)*. Singapore: Curriculum Planning Division.
- Singapore Ministry of Education (2004a). *Assessment guides to primary mathematics*. Singapore: Author.
- Singapore Ministry of Education (2004b). *Assessment guides to lower secondary mathematics*. Singapore: Author.

- Stenmark, J. (1991). *Mathematics assessment: Myths, models, good questions, and practical suggestions*. Reston, VA: National Council of Teachers of Mathematics.
- Wisconsin Education Association Council. (1996). *Performance assessment*. Madison, WI: Education Issues Series.
- Wong, N.-Y., Lam, C.-C., Wong, K.-M. P., Leung, F. K.-S., & Mok, I. A.-C. (2001). Students' views of mathematics learning: A cross-sectional survey in Hong Kong. *Education Journal*, 29(2), 37-59.
- Wu, H. (1994). The role of open-ended problems in mathematics education. *Journal of Mathematics Behavior*, 13(1), 115-128.

Appendix

Students' Attitude toward Mathematics and Mathematics Learning Survey Items

A. Part I (for both Pre- and Post-intervention surveys)

1. I enjoy doing mathematics.
2. I am never under a terrible strain in a math class.
3. I am sure I can learn mathematics well.
4. I believe mathematics is useful.
5. Mathematics is hard for me.
6. I am not afraid of doing mathematics.
7. I can get good grades in mathematics.
8. It is important to know mathematics nowadays.
9. Mathematics is interesting to me.
10. I am unable to think early when doing mathematics.
11. I am not good at mathematics.
12. Studying mathematics is a waste of time.
13. I don't have good feelings about mathematics.
14. I feel lost when trying to solve math problems.
15. I don't think I can do well in mathematics.
16. I like spending time on studying mathematics.
17. It makes me nervous to even think about having to do a math problem.
18. I will use mathematics a lot as an adult.
19. I don't like to attend math lessons.
20. I have a lot of confidence when it comes to mathematics.
21. I like solving challenging math problems.
22. I would rather have someone give me the solution to a difficult math problem than to have to work it out for myself.

B. Part II (for pre-intervention survey; only items relevant to performance tasks are included)

24. In the last school term, my math teacher encouraged me to solve math questions in different ways.
27. In the last school term, how many math questions did your teacher ask you to do that have more than 1 correct answer?

28. In the last school term, how many math questions did your teacher ask you to do that have nothing to do with real life situations?

C. Part II (for post-intervention survey)

23. I like to solve mathematics questions which have more than one correct answer.
24. Doing mathematics performance tasks is difficult to me.
25. Doing performance tasks helps me to learn mathematics.
26. I like to do mathematics questions which could be solved using different methods.
27. Doing mathematics performance tasks help me to be more creative in problem solving.
28. I have to think harder when I am doing mathematics performance tasks.
29. I feel lost when I am doing mathematics performance tasks.
30. I like to do mathematics questions which involve the real world.
31. Doing mathematics performance tasks helps me see more connection between mathematics and daily life.
32. Doing mathematics performance tasks helps me to become more systematic when I am solving mathematics problems.
33. I need hints to help me do mathematics performance tasks.
34. I am good at doing mathematics performance tasks.
35. Doing performance tasks takes me more time than doing other mathematics questions usually done in class.
36. I would like to have more mathematics performance tasks for my mathematics lessons.
37. Doing mathematics performance tasks makes me learn mathematics better.
38. Doing mathematics performance tasks is a waste of time.

Authors:

Lianghuo Fan

Nanyang Technological University, Singapore

Email: lianghuo.fan@nie.edu.sg

Yan Zhu

University of Hong Kong, Hong Kong

Email: yanzhu@hkucc.hku.hk