Pilot Study on the Validity and Reliability of MIM: An Alternative Assessment for Measuring Metacognition in Mathematics among College Students

Chak Him FUNG

Chi Keung LEUNG

Department of Mathematics & Information Technology
Education University of Hong Kong
Tai Po, New Territories
Hong Kong

Abstract

In this study, an alternative assessment for measuring metacognition in mathematics (MIM) is proposed. The reliability of the MIM is examined (Cronbach α=.905) while the Split-half reliability test is done (Cronbach α = .825). The factor analysis yielded four factors which were 'prediction', 'planning', 'monitoring' and 'evaluation'. Meanwhile, the MIM yielded a small linear relationship with the High School Entrance Exam (HSEE) (r=.285, ρ<.001) which is highly consistent with recent research findings. Planning seems to have the highest linear relationship with mathematics performance among all metacognitive skills (r=.346, ρ<.001). On the other hand, the academic performance of senior two students (M=76,SD=22.2) was found to be significantly lower than those of senior three students (M=170,SD=19.0) while no significant difference was found in metacognition. It is believed that the Content-specific knowledge would become more and more dominant in predicting the academic success when the level of study increases.

Keywords: Metacognition, alternative assessment, domain-specific instrument and mathematical modelling.

Introduction

1.1 Definition of Metacognition

Metacognition was first introduced by Flavell (1976) to explain the difference in the learning strategies used by students in different age. The word “meta” means beyond thus metacognition is always considered as “the thinking about thinking”.

I am engaging in metacognition if I notice that I am having more trouble learning A than B; if it strikes me that I should double-check C before accepting it as a fact; (...) if I become aware that I am not sure what the experimenter really wants me to do; if I sense I had better make a note of D because I may forget it; if I think to ask someone about E to see if I have it right. (Flavell, 1976, p. 232)

He further defined metacognition as the knowledge and active monitoring of one's own cognitive processes. Metacognition can be distinguished as metacognitive knowledge and metacognitive regulations or skills (Flavell, 1979, 1987). According to Snowman and McCown (2012), there are three kinds of metacognitive knowledge: (1) Knowledge of person, which refers to the knowing of the strength and weakness about himself, (2) Knowledge of task, which means the knowing about the difficulties of the task, and (3) Knowledge of strategy, which refers to the knowing about suitable methods in solving the problem. According to Brown (1978, 1987),metacognitive skills can be divided into (a) Prediction (eg. How difficult is the task), (b) Planning (eg. What shall I do to execute the task), (c) Monitoring (eg. What do I yet not know in order to attain my objective) and (d) Evaluation (eg. Have I got the full meaning of the answer) as shown in Figure 1.

Meanwhile, metacognitive knowledge and skills interact with each other. Metacognitive knowledge can be added, deleted, or changed through metacognitive regulation (Flavell, 1979, p. 906).According to the nature of the metacognitive skills, some researcher further classified Prediction and Evaluation as offline metacognitive skills because they are measured before or after the solving of exercises. In contrast, Planning and Monitoring are classified as on-line metacognitive skills (Desoete & Roeyers, 2002).

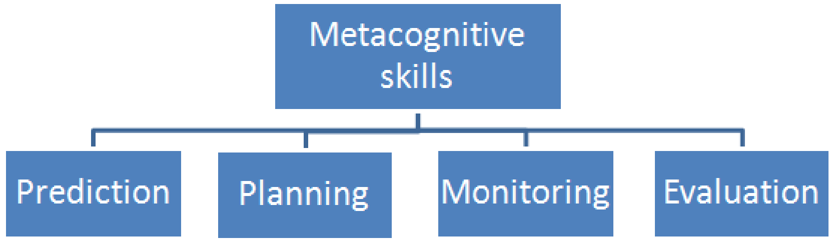

Figure1. The Four Types of Metacognitive Skills

1.2 Significance of Metacognition

Educational psychologists have long promoted the importance of metacognition for regulating and supporting student learning (Lai, 2011& Sperling, Howard, Miller& Murphy, 2002). Research shows that there is a significant relationship between the metacognitive skills and the mathematical competence among children (Desoete & Roeyers, 2002). As Desoete & Veenman (2006) state, "metacognition seems involved, especially during the initial stage of mathematical problem solving when students build an appropriate representation of the problem and plan their problem-solving steps, as well as in the final stage of interpretation and checking the outcome of calculations." It would prevent 'blind calculation' or a superficial 'number crunching' approach. Metacognition could reflect students' true mathematical abilities.

On the other hand, metacognition could enhance effective teaching. Hart & Memnum (2015) report that teachers who have a higher level of metacognitive awareness teach more effectively. Mathematics teaching-Learning process could be improved if teachers would spend more time on metacognitive instruction (Desoete 2007). If textbooks and homework are enriched with metacognitive strategies, students' academic results could be improved (Mesa, 2004; Chan & Leung, 2013; Özcan & Erktin, 2015; Mandaci Sahin & Kendir, 2013). Although some researchers argue that metacognition does not significantly correlate with students'' GPA (Cetin, 2017), metacognition is believed to be significant in students' mathematics performance by the majority (Pintrich & De Groot, 1990; O'Neil Jr. & Abedi, 1996; Sperling, Richmond, Ramsay, & Klapp, 2012; Özsoy, 2011).

1.3 Challenges in Measuring Metacognition

Assessment of metacognition is challenging because they are a part of mental processes of the individual (Favieri, 2013). Although metacognition could enhance students' academic performances and achievements, standardized achievement scores such as GPA are not good indicators for metacognition. Students' mathematical abilities are not revealed thoroughly by the traditional test (Stephens, 1998). At the same time, as Sperling et al., (2002) state, "research indicates that the relationship between standardized academic achievement scores and metacognition is not direct." The relationship between students' standardized achievement scores and metacognition is still unclear. Desoete & Veenman (2006) suggest that some metacognitive skills seem to be more related to mathematical school success than others. In more detail, "Planning" has a higher significance than control skills in predicting a better score in grade 5 students. Although the goal setting, planning and control skills are generally correlated, they can be developed separately. Standardized academic achievement, however, can be problematic when used in measuring metacognition (Sperling et al., 2002).

Lai (2011) proposes four main reasons in explaining the difficulties: "(1) metacognition is a complex construct, involving a number of different types of knowledge and skills; (2) it is not directly observable; (3) it may be confounded in practice with both verbal ability and working memory capacity; and (4) existing measures tend to be narrow in focus and decontextualized from in-school learning."

1.4 The Need for a Quantitative Metacognitive Instrument Specific to Mathematics

In order to assess metacognition among students better, different measuring instruments with a large variety of measuring method were proposed. Instruments such as questionnaires, interviews, observations, thinking-aloud protocols, eye movements, computer registrations of activities, note taking, stimulated recalls have been used (Desoete & Veenman, 2006). Some researchers suggest qualitative measurements, such as group interview (Dahl, 2004; Swanson, 1990) while some suggest quantitative measurements, such as self-assessed questionnaire (Isquith & Gioia, 2008). It is believed that no one single assessment is the best. A method may be more suitable for measuring a particular metacognitive component while another method fits better in other components (Veenman, 2005). Sperling et al. (2002) suggest that each method of measurement have its own strengths and weaknesses.

For example, results collected from oral interviews are being doubted (McLain, Gridley & McIntosh, 1991). Although it could externalize participants' thought there could be a gap between participants' conversations and actions, especially among children (Brown, 1980). Among them all, self-report inventories are believed to be a better measuring instrument in the metacognitive process with the least problematic technique (Sperling et al., 2002). Quantitative measures are more efficiently administered and scored than qualitative measures such as think-aloud protocols (Everson & Tobias, 1998). With regard to the study made by Sperling et al. (2002), quantitative measures in students' metacognition play a significant role in investigating the relationships between their academic performance and metacognition.

Moreover, the possibility and potential of being used as an alternative assessment are also considered. Assessment in mathematics has traditionally been measurement-driven because the assessment is expected to carry both ranking and screening effect. It also has to keep the accountability of the educational system (Broadfoot, 1996). Thus traditional mathematics tests focus on repeating the learning process by using only limited types of questions (Firestone,Winter& Fitz, 2000). It is doubted that the tests could reveal students' true abilities completely. However, teachers always feel reluctant towards the using of alternative assessments such as observation, interviews, demonstrations and practical investigations. Watt (2005) concluded that the main objection to alternative assessment methods was that they were too time-consuming or too subjective. Although traditional mathematics tests are not the best instrument for examining students' mathematical abilities, it is still very popular. In view of that, a quantitative instrument for measuring metacognition could serve as an alternative measure to the traditional mathematics tests.

On the other hand, the domain-general/domain-specific issue in metacognition remains unsolved (Veenman, van Hout-Wolters & Afflerbach, 2006). Some researchers suggest that metacognition is domain-general since there are some general strategies or skills among domains (Schraw, Dunkle, Bendixen & Roedel, 1995). An improvement in such strategies or skills could enhance the learning in all domains. In contrast, some researchers report that students would apply different metacognitive strategies for different subjects (van der Stel & Veenman, 2008). Other study shows that metacognition could be partially domain-specific (Wang, 2015). All of the above researchers provide experimental data in supporting their claims and it is difficult to make a conclusion at this instant. However, authors of this paper tend to believe that metacognition is at least partially domain-specific. Variations in metacognition could be found across different domains between subjects or Key Learning Areas such as mathematics and English reading comprehension. It is because some of the metacognitive skills might be more useful in some particular subjects but not in others. For example, alternative method is one of the common strategies used in mathematics, especially in algebra. After yielding "$x = 3$" by subtracting both sides by 2 from "$x + 2 = 5$", an addition of "$3+2$" is done for checking the correctness of the answer. Relatively, this strategy plays a less important role in English writing. On the other hand, skimming is one of the metacognitive reading skills (Hartman, 2001). An improvement in skimming skills would enhance academic performance in reading dramatically. However, this strategy seems to be less significant in solving mathematical questions. Even though there may exist some generic metacognitive skills among subjects, one may be good at using suitable mathematical strategies or equations in solving mathematics problems but not as good as he does in reading comprehension (Desoete & Roeyers, 2002). As a result, the domain-specific characteristic of metacognition is significant. Validated instruments designed for measuring the metacognition in mathematics are very limited in the field (Favieri, 2013). Thus a quantitative tool for measuring metacognition in mathematics is more preferred especially for alternative assessment in purpose.

1.5 The Existing Quantitative Tools

In order to measure the learning process, different researchers choose different quantitative tools on purpose. For example, tools such as The Behavior Rating Inventory of Executive Function (BRIEF) (Isquith & Gioia, 2008), Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, Smith, Garcia & McKeachie, 1991), the Learning and Study Strategies Inventory (LASSI) (Weinstein, Palmer & Schulte, 1987; Eldredge, 1990), Junior Metacognitive Awareness Inventory (Jr.MAI) (Sperling et al., 2002) and The Metacognitive Awareness Inventory (MAI) (Schraw & Dennison, 1994) are widely used. According to the authors, BRIEF is designed to figure out the relationship between behaviour and cognitive process (Pintrich et al., 1991). LASSI is designed to measure learning strategies and attitude (Eldredge, 1990), while MSLQ mainly focuses on motivation and learning strategies (Pintrich et al., 1991).

These inventories all contain some (but not a complete set of) components of metacognition. They are not very suitable for being a measurement of metacognition. MAI and Jr. MAI could be considered as two better instruments in measuring the metacognition. Jr. MAI is a modified version of MAI. MAI is a self-rated assessment which contains 52 items while Jr. MAI contains 18. MAI and Jr. MAI aim to measure the metacognitive awareness of university students (Schraw & Dennison, 1994) and the metacognitive awareness of children at grades 3 to 9 level (Sperling et al., 2002) respectively. However, they could be considered as general metacognitive instruments rather than domain-specific instruments to mathematics. In other words, they are good at measuring students' general metacognition among subjects but not suitable for a specific one such as mathematics. Favieri (2013) thus proposed general Metacognitive Strategies Inventory (GMSI) and Metacognitive Integrals Strategies Inventory (MISI), which were modified from the MAI, in order to suit college students. Similar to MAI, GMSI measures the students' general metacognition. However, MISI measures students' metacognition in integration only. It is considered as a task-specific metacognition instrument instead of a domain-specific instrument.

Meanwhile, an important component "Prediction" is missed among them. Study shows that the prediction is significantly lower among children with mathematics disable compared with age-matched peers (Desoete & Roeyers, 2002). According to Brown (1978), prediction requires the abilities to imagine cognitive acts that have not yet occurred. It enables students to think about the available time, learning characteristics and learning objects (Desoete & Roeyers, 2002). The estimation or prediction of difficulty of a task would regulate the outcome and efficacy expectation (Winne, 1997). It allows students to spend more effort on difficult tasks but lesser on easier tasks. Leung (2011) suggests that through the questionings about students' decision in determining the difficulties among tasks, declarative and conditional metacognitive knowledge would be enhanced. If a pattern could be observed by the students, at last, their metacognitive knowledge and skills improved. Thus, prediction should be measured as a component in metacognition.

Moreover, latest research trends are shifting its focus to investigating the degree of complexity and mechanism in metacognition by comparing it by other cognitive models rather than using metacognition as the single independent variable alone. For example, the research conducted by Bryce, Whitebread & Szűcs (2015) used two assessment tools (one for metacognition and the other for executive function) in order to measure the relationship among executive functions, metacognitive skills and educational achievement. Although having very high validity and reliability, a long assessment with numerous items is not preferred in such cases. Let's say if the target sample could bear an assessment with a maximum of 40 items (item capacity of 40) and more than one assessment is planned to be taken, it is irrational to occupy the whole capacity with one single assessment. Although the precision of one assessment decreases, room should be left in order to maintain the precision of other assessments.

1.6 Design of the MIM (Stage A): Self-Assessment

With regard to the domain-specific property of metacognition and the importance of prediction, a new quantitative instrument called metacognitive inventory for mathematics (MIM) was introduced. Since metacognitive knowledge and regulation are related to each other (Pintrich, & De Groot, 1990; Schraw, 2001; Schraw, & Dennison, 1994; Swanson, 1990), the quantitative data yielded from either of them could still have significant representativeness to metacognition. Since complexity makes unreliability an issue (Schraw and Moshman, 1995), MIM measures only the four metacognitive skills in metacognition in order to maximize its validity and reliability. The subsection of MIM will follow the metacognitive skills model suggested in Figure 1. With reference to MAI and Jr. MAI, it is assumed that the full item capacity of university students and grade 9 students are 52 and 18 respectively. By linear assumption, senior secondary school students could have an item capacity of $18+(52-18)/(13-9)\times(11-9)=35$, of which university students and senior students are treated as grade 13 and 11 in the calculation. Thus the length of MIM is limited to 16 items in order to balance the problem of precision and item capacity. The construction of the items follows the Brown's model in metacognitive skills (Brown 1978, 1987) whereas it is domain-specific to mathematics. In detail, "Prediction" includes the objective, difficulty, time and effort estimation. "Planning" includes determining suitable or useful information, mathematical equations, strategies and procedures. "Monitoring" refers to the checking of time, plans, strategies and procedures while "Reflection" means the reflective thinking in the micro & macro meanings of the work, elaborations and efficiency. The word "mathematical" appears several times not only in the title but also the items in MIM. Vocabularies such as "plans" and "strategies" generally refer to all plans and strategies within the mathematics domain. This provides the face validity of MIM.

Only four items are allocated under each subsection in order to keep college students' focus. In short, MIM focuses mainly on both online and offline skills. It aims at producing acceptable validity and reliability with the least items. The MIM follows the 7-points Likert Scale so that the sensitivity of measuring the degree of strength in each item could be maximized. Students are free to choose between 1 to 7 in each item. "1" represents the lowest degree while "7" represents the highest or greatest degree in each item. "0" is omitted because it may mislead students by implying "nothing". All items are compulsory. Each student will have to fill in 16 items in total. It is expected that each student spends 30 seconds for each item, whereas the time spent in filling in the report is about 8 minutes. Since the mother language of the target sample in this study is Chinese, translation of the MIM will also be given beside each item in order to make effective communication. Difficult terms are avoided since it may reduce the reliability (O'Neil & Abedi, 1996). The score of the MIM is calculated by summing the average of all items under the subsection with identical weightings.

1.7 Design of the MIM (Stage B): Peer-Assessment

Obviously, the problem of subjectivity could also be questioned in MIM since it is also classified as a self-rated assessment. The metacognitive inventory for mathematics (MIM) consists of peer-assessment is then suggested. The relationship between the score in self-rated and the peer-rated MIM is studied. It is believed that a more precise view of an individual's metacognitive abilitiescouldbe obtained from information derived from the response of participants (Saldaña, 2004). Experience has shown that groups of three to five students work well (Posamentier & Stepelman, 1999). It is expected that students are groupedintofouraccording to this rule of thumb and thus every student is supposed to fill in three peer-assessments. According to the nature of peer-assessment, the prerequisite is that the assessors must have a basic understandingabout the subject. The peer-assessment will be done after a game-like activity in this study. It will serve as the basis for the peer-assessment. Moreover, in order to simplify the assessment, only one single question will be asked in each subsection.

2. Research Objective

The objective of the research includes the followings:

1. To introduce MIM in assessing metacognition in mathematics.
2. To study if metacognition could predict academic performance by examining the total percent of variance explained.
3. To study the relationship between the scores in self-rated and the peer-rated MIM.

3.Methodology

3.1 Research Flow

The research is divided into two stages. First, the validity and the reliability of the MIM are verified. A game-like activity is carried out and the relationship between the scores in the self-rated and the peer-rated MIM is determined. A follow-up individual interview is made in order to investigate other possible factors which may affect the result of this study.

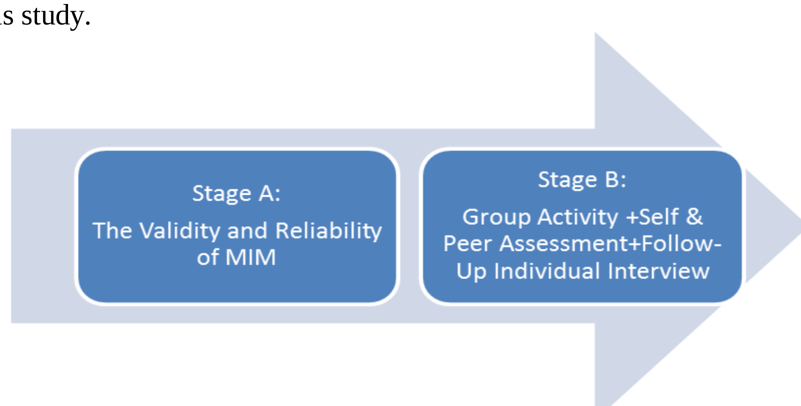

Figure 2. The Study Flow

3.2 Participants

1100 senior one students of a public secondary school are enrolled in stage A while 14 senior two and 14 senior three voluntary students studying the international program are enrolled in stage B. All 28 students had taken the same GCE AS level Unit C1&2 Math Examination (AS C12) in June 2016.

3.3 Practical Procedure

In stage A, the MIM was distributed to 1100 senior one students on 20 Oct 2016. The validity and reliability of the MIM were verified. Mathematics score in the high school entrance exam (HSEE) was used as the indicator of mathematics performance. The HSEE is an integrated uniform test examining for a wide range of mathematical knowledge such as algebra, statistics and probability etc. It is believed that such a uniform test could have a high degree of representativeness for mathematics performance. Since it is a systematic examination provided by the educational council, it is considered as a least problematic indicator.

In stage B, 28 voluntary students of senior 2 and 3 were then engaged in a group activity on 28 Oct 2016. The game-like group activity was carried out with a problem-solving approach. It took about 20 minutes. Each group of students engaged in two problems and each member of the groups got different card sets which carried some necessary pieces of information. No member of a group shared the same set of information. Meanwhile, no member got all necessary pieces of information and thus group work was needed in order to solve the problems. All answers were to be written on the answer sheet provided. However, these answers itself would not be scored in the MIM because collaborating metacognition was not the focus of this research. The MIM containing both self-assessed and peer-assessed parts were given to them. The correlation between the self-rated score and the peer-rated score was found.

Their British GCE AS Examination score in Math Unit C1&2 (AS C12) was taken into analysis next as the indicator of their mathematics performance. Similarly, the British GCE examination is an international public examination for the entrance of university. It exams a wide range of mathematical knowledge. It is believed to have a high degree of representativeness. All data analysis was done using SPSS 24.

At last, an individual interview of four randomly selected students was done in order to figure out other possible factors which would account for their mathematics performance.

3.4 Data Collection and Handling of Data

All of the MIM questionnaires were collected. Among them, one questionnaire collected in stage A was rejected since the data reported was considered as invalid. In that data, the reported HSEE was 8. However, the minimum entrance mark of this high school in the academic year (2016-2017) was 416 while the full mark of this exam is 530. Mathematics accounts for 120 marks out of the total. An “8” in mathematics was very likely to be an outlier and thus is rejected.

14 senior two and 14 senior three students took part in the second stage of the experiment as volunteers. Two MIMs in senior 2 contained missing parts in the peer-assessment and thus were eliminated from the analysis. Demographic details were not collected because the samples were assumed to be homogenous in terms of characteristics such as age range.

4. Result, Analysis and Discussion

4.1 Internal Consistency Reliability

Table 1. Cronbach's value of items in MIM

Item-Total Statistics of the MIM	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Self-rated item Q1	65.71	251.415	.551	.900
Self-rated item Q2	66.21	248.715	.575	.900
Self-rated item Q3	65.32	255.084	.490	.902
Self-rated item Q4	65.63	247.410	.647	.897
Self-rated item Q5	65.59	251.369	.556	.900
Self-rated item Q6	65.50	250.250	.620	.898
Self-rated item Q7	66.41	246.493	.592	.899
Self-rated item Q8	65.98	247.405	.649	.897
Self-rated item Q9	66.65	247.759	.611	.898
Self-rated item Q10	66.17	246.806	.587	.899
Self-rated item Q11	66.48	246.456	.601	.899
Self-rated item Q12	66.00	246.483	.560	.900
Self-rated item Q13	65.61	249.941	.573	.900
Self-rated item Q14	65.86	249.319	.568	.900
Self-rated item Q15	65.70	248.415	.580	.899
Self-rated item Q16	66.12	248.627	.533	.901

The 16 self-rated items in the MIM were found to be highly reliable (α=.905). Table 1 shows all items are significant in measuring a single criterion and none of them should be deleted. The MIM which consists of both self-assessed and peer-assessed items (N=28) proved to have high reliability (α=.908). It shows that the 28-item with peer-assessment is also consistent in measuring a single criterion.

4.2 Split-Half Reliability

The MIM was then put into a split-half test by selecting two items per each subsection. In practice, the split-half test was done by separating the odd items from the even items. The MIM shows high internal consistency in the split-half reliability test (α=.825).

4.3 Construct Validity

Table 2. Pattern Matrix of the MIM in Factor Analysis

Pattern Matrix[a]	Component 1	2	3	4
Self-rated item Q6	.800			
Self-rated item Q8	.689			
Self-rated item Q7	.634			
Self-rated item Q5	.626			
Self-rated item Q12	.526		-.304	
Self-rated item Q14		.842		
Self-rated item Q15		.813		
Self-rated item Q16		.687		
Self-rated item Q13		.546		
Self-rated item Q10			-.837	
Self-rated item Q11			-.812	
Self-rated item Q9			-.746	
Self-rated item Q3				.800
Self-rated item Q1				.672
Self-rated item Q2				.558
Self-rated item Q4				.523

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 10 iterations.

The values in KMO and Bartlett's Test were significant (KMO=.930, χ^2 =7046.4, ρ<.001). Factor analysis with Oblimin rotation was performed. The best solution yielded four factors with 62.7% of the variances among the items could be explained by them. Items and their corresponding factor were shown in Table 2. The four factors were classified as 'prediction', 'planning', 'monitoring' and 'evaluation' according to their loadings, which were consistent with the model of metacognition suggested. The MIM was believed to have high construct validity.

4.4 CriterionValidity and External Validity

The MIM was found to have a small linear relationship with the HSEE (r=.285, ρ<.001). It provided the grounding evidence for the criterion validity of the MIM. In similar research, Sperling et al. (2012) report this coefficient as r=.26 with ρ<.01. Pintrich and De Groot (1990) report that this correlation as a range from .07 to .36 using the MSLQ. Özcan (2014) yields .42 and .26 with ρ<.001 and ρ<.01 respectively. Thus, although small, the result in this study is consistent with studies done by other researchers.

Linear regression analysis was carried out by using SPSS 24. The prediction, planning, monitoring and evaluation together could significantly predict one's performance in HSEE with F(4,1094) = 38.82, ρ<.001. The four predictors, together, could explain 12.4% of the total variance of the HSEE performance. The planning (β=.334, t=9.797, ρ<0.001) and evaluation (β=.066, t=1.960, ρ=.050) were significant predictors of the HSEE performance whereas prediction (β=.017, t=.537, ρ=.591) and monitoring (β=.053,t=1.586, ρ=.113) did not significantly predict the HSEE performance.

Meanwhile "planning" yielded the highest linear correlation with the HSEE among all metacognitive skills (r=.346, ρ<.001). Mytkowicz, Goss, & Steinberg (2014) report a similar correlation coefficient with r=.385 (ρ= .009). It is also consistent with the findings suggested by the previous study done by Desoete & Veenman (2006), who report that "planning" seemed to have the highest linear correlation with academic performance among metacognitive skills.

4.5 Correlation between Self-rated and Peer-rated items in MIM

Results revealed that there is a strong correlation between the self-rated and peer-rated items in the MIM (r=.509, ρ= .008). Such strong linear relationship shows that the MIM is still valid although it might be subjective.

5. Further Discussion

5.1 Effect of Metacognition and the Years of Study on Mathematics Performance

It is generally believed that higher education will yield a higher metacognitive score. The mean AS C12 score of senior two students (M=76,SD=22.2) was found to be significantly lower than those of senior three students (M=170,SD=19.0) as expected (F(1,24)=136.403, ρ<.001). However, it was surprising that the main effect of metacognition was not significant in AS C12 score with F(1,22)=.517,ρ=.480.At the same time, the interaction effect of metacognition was not significant either(F(1,21)=2.009,ρ=.171). It suggested that the metacognition among senior three students did not significantly differ from those of senior two students. Linear regression model shows that the year of study alone could actually explain 84.3% of the total variance of the AS C12 score.

One of the possible reasons could be the lack of metacognitive training in the last academic year for the present senior three students. However, this alone was not satisfactory in explaining the huge differences in average scores. Meanwhile, Sperling et al. (2002) report that the correlation between metacognition and academic success decreases in older population. Bryce et al. (2015) found similar results among grade 5 and grade 7 students. It is more likely that the knowledge in each subject taught becomes more content-specific when learners' are getting older. A follow-up individual interview was conducted and interviewees reported similar results. Students of senior 2 claimed that the reason for their poor performance is mainly due to the incompletion of their AS course taught. The knowledge, formula and methodologies are nearly impossible for them to understand or even imagine before the lessons. The content-specific knowledge (CSK) would become more and more dominant in predicting the academic success when the level of study increases. Together with the results above, the predictive power of the domain-specific metacognition is weakened although it is still valid. Further research is suggested in order to understand the relationship between factors in CSK and metacognition.

5.2 Non-linear Property of Monitoring

Monitoring is found to have a slightly negative linear relationship with the HSEE in this study with r = -.129, ρ<.001. Similar results are also reported by Bryce et al. (2015). Further investigation in curve estimation is then carried out and the results are shown in Table 3.

Table 3. The Model Summary of Monitoring

Model Summary and Parameter Estimates									
Dependent Variable: High School Entrance Exam									
Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	.037	42.290	1	1097	.000	97.528	.320		
Logarithmic	.027	30.706	1	1097	.000	92.282	3.831		
Inverse	.015	16.355	1	1097	.000	104.894	-30.903		
Quadratic	.043	24.723	2	1096	.000	102.053	-.307	.019	
Cubic	.045	17.077	3	1095	.000	106.637	-1.364	.091	-.001
Compound	.035	40.341	1	1097	.000	97.085	1.003		
Power	.026	29.352	1	1097	.000	92.071	.039		
S	.014	15.527	1	1097	.000	4.650	-.311		
Growth	.035	40.341	1	1097	.000	4.576	.003		
Exponential	.035	40.341	1	1097	.000	97.085	.003		
The independent variable is Self-rated Monitoring.									

Interestingly, a quadratic (R^2=.43,F(2,1096)=24.7,ρ<.001) or a cubic model (R^2=.45,F(3,1095)=17.1,ρ<.001) could explain more variance of the HSEE than the linear model (R^2=.37,F(1,1097)=42.3,ρ<.001). Both models suggest that there is a trough in the middle of the HSEE. It further implies that the academic performance will drop first and rise again after the monitoring increases beyond a certain level. It is believed that the proficiency of the monitoring skill is one of the main causes. In more detail, one of the main functions of the HSEE is to provide a screening effect such that the score in the examination could be used as an indicator in selecting the best student with the greatest academic performance. Here, the best could refer to those who could finish the questions quickly and accurately within the time limit. For those who check themselves frequently is believed to have mastered the skills of monitoring. They would achieve the exam with a high accuracy and a good time management by proficiency. On the other hand, students seldom check themselves have an advantage in time management. However, the increase of the accuracy is balanced by the decrease of the accuracy for students whose have an average monitoring scores. As a result, a trough is formed. Further study is suggested to investigate a more accurate model of metacognition.

Limitation

The loadings in Table 5 show that item Q12 measures both the planning and the monitoring. However, it seems to measure the planning more. The main difference between Q7 and Q12 is that Q7 is a process during "planning" while Q12 is the one after it. These two statements might be confusing. Improvement is made in bolding the word "…after my calculation." in Q12. The wording is also moved to the front of the sentence. The MIM was conducted in a single school only. At the time of the present study, the participating school was ranked no.3 in the district and thus there might be a ceiling effect or a cutting off effect in the HSEE marks. The spreading out of the mark of the exam is thus narrow. It is believed that the *r* value could be higher if more schools in different banding could take part in this research.

Conclusion

In this research, the reliability of the MIM was examined (Cronbach α=0.905). Split-half reliability test was done and the Cronbach α value is .825. The factor analysis with Oblimin rotation yielded four factors which were classified as 'prediction', 'planning', 'monitoring' and 'evaluation' according to the content of the items. In criterion validity and external validity test, the MIM yielded a small linear relationship with the HSEE (r=.285, ρ<.001) which is highly consistent with recent research findings. Planning seems to have the highest linear relationship with mathematics performance among all metacognitive skills (r=.346, ρ<.001).

At the same time, the self-rated items are found to have a significant relationship with the peer-rated items (r=.509, ρ= .008). It suggests that the MIM is a valid instrument although it is a self-rated instrument. Meanwhile, it is found that the metacognition could determine about 12.5% of the total variance of the HSEE. It further suggests that metacognition is a significant predictor to the HSEE.

On the other hand, the academic performance of senior two students (M=76,SD=22.2) was found to be significantly lower than those of senior three students (M=170,SD=19.0) with no significant difference in metacognition. It is believed that the CSK would become more and more dominant in predicting the academic success when the level of study increases. Moreover, a quadratic (R^2=.43,F(2,1096)=24.7,ρ<.001) or a cubic model (R^2=.45,F(3,1095)=17.1,ρ<.001) could explain more variance of the HSEE than the linear model (R^2=.37,F(1,1097)=42.3,ρ<.001). It is believed that the proficiency of the monitoring skill is one of the main causes.

In conclusion, the MIM is a reliable and valid instrument for measuring the metacognition among all its components which are prediction, planning, monitoring and evaluation. It could be an effective alternative assessment for measuring the true mathematics ability among students.

Acknowledgements

The authors would like to thank Professor Joyce, L.Y. Kwan as a consultant for helpful comments and suggestions for the statistical analysis in this paper.

Hong Kong Mathematics Education Conference

This paper was partly presented at the Hong Kong Mathematics Education Conference in May 2017.

References

梁志強(2011)：小學生數學後設認知的培養，《數學教育期刊》45，85-94。

陳小蘭和梁志強（2013）：比較香港和新加坡小學數學課程中後設認知的培養，《數學教育期刊》，46，40-59。

Broadfoot, P. (1996). Education, Assessment and Society. Buckingham, Open University Press.

Brown, A. L. (1978). Knowing when, where, and how to remember. A problem of metacognition. In R. Glaser (Ed.), Advances in instructional psychology: volume 1 (pp.77-165). Hillsdale, NJ: Erlbaum.

Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiror, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 453-481). Hillsdale, NJ: Lawrence Erlbaum.

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F.Weinert & R. Kluwe, (eds.), Metacognition, Motivation, and Understanding (pp. 65–116). Hillsdale, NJ: Erlbaum.

Bryce, D., Whitebread, D., & Szűcs, D. (2015).The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children.Metacognition and Learning,10(2), 181-198.

Cetin, B. (2017), Metacognition and Self-regulated Learning in Predicting University Students' Academic Achievement in Turkey.Journal of Education and Training Studies, 5(4).

Dahl, B. (2004).Analysing cognitive learning processes through group interviews of successful high school pupils: development and use of a model.Educational Studies in Mathematics,56(2), 129-155.

Desoete, A. (2007). Evaluating and improving the mathematics teaching-learning process through metacognition. Electronic Journal of Research in Educational Psychology,5(3), 705-730.

Desoete, A.& Roeyers, H. (2002). Off-line metacognition — a domain-specific retardation in young children with learning disabilities?.Learning Disability Quarterly,25(2), 123-139.

Desoete, A. & Veenman, M. (2006). Metacognition in mathematics: Critical issues on nature, theory, assessment and treatment. In A. Desoete & M. Veenman (Eds.), Metacognition in mathematics education (pp. 1–10).Haupauge, NY: Nova Science.

Eldredge, J. L. (1990). Learning and study strategies inventory - high school version (lassi - hs). Journal of Reading,34(2), 146-149.

Everson, H. T., & Tobias, S. (1998). The ability to estimate knowledge and performance in college: a metacognitive analysis. Instructional Science,26(1), 65-79.

Favieri, A. G. (2013). General metacognitive strategies inventory (GMSI) and the metacognitive integrals strategies inventory (MISI). Electronic Journal of Research in Educational Psychology,11(3), 831-850.

Firestone, W. A.,Winter, J., & Fitz, J. (2000). Different assessments, common practice?mathematics testing and teaching in the USA and England and Wales. Assessment in Education: Principles, Policy & Practice,7(1), 13-37.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), The nature of intelligence (pp. 231-236). Hillsdale, NJ: Erlbaum.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34, 906-911.

Flavell, J. H. (1987). Speculations about the nature and development of metacognition.In F. E.Weinert & R. H. Kluwe (Eds.), Metacognition, Motivation and Understanding (pp. 21-29).Hillside, New Jersey: Lawrence Erlbaum Associates.

Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills.In H. J. Hartman (Ed.), Metacognition in learning and instruction (pp. 33–67). Dordrecht, Netherlands: Kluwer.

Hart, L. C., & Memnun, D. S. (2015).The relationship between preservice elementary mathematics teachers' beliefs and metacognitive awareness. Journal of Education & Training Studies,3, 70-77.

Isquith, P. K., & Gioia, G. A. (2008). Behavior rating inventory of executive function. PAR, Odessa.

Lai, E. R. (2011). Metacognition: A Literature Review. Pearson Research Report. Retrieved from *https://psychcorp.pearsonassessments.com/hai/images/tmrs/Metacognition_Literature_Review_Final.pdf*.

Mandaci Sahin, S., & Kendir, F. (2013).The effect of using metacognitive strategies for solving geometry problems on students' achievement and attitude. Educational Research & Reviews,8, 1777-1792.

Mclain, K. V. M., & Mcintosh, D. (1991). Value of a scale used to measure metacognitive reading awareness. Journal of Educational Research,85(2), 81-87.

Mesa, V. (2004). Characterizing practices associated with functions in middle school textbooks: an empirical approach. Educational Studies in Mathematics,56(2), 255-286.

Mytkowicz, P., Goss, D., & Steinberg, B. (2014). Assessing metacognition as a learning outcome in a postsecondary strategic learning course. Journal of Postsecondary Education & Disability,27, 51-62.

O'Neil Jr., H. F.& Abedi, J. (1996). Reliability and validity of a state metacognitive inventory: potential for alternative assessment. Journal of Educational Research,89(4), 234-245.

Özcan, Z. Ç. (2014). Assessment of metacognition in mathematics: which one of two methods is a better predictor of mathematics achievement? International Online Journal of Educational Sciences,6(1), 49-57.

Özcan, Z. Ç., Erktin, E. (2015). Enhancing mathematics achievement of elementary school students through homework assignments enriched with metacognitive questions. Eurasia Journal of Mathematics Science & Technology Education,11(6), 1415-1427.

Özsoy, G. (2011). An investigation of the relationship between metacognition and mathematics achievement. Asia Pacific Education Review,12(2), 227-235.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology,82(1), 33-40.

Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire. Ann Arbor, MI: The Regents of The University of Michigan Technical Report 91-B-004.

Posamentier, A. S. & Stepelman, J. (1999). Teaching Secondary School Mathematics: Techniques and Enrichment Units (5thed.). New Jersey: Prentice Hall.

Saldaña, D. (2004). Interactive assessment of metacognition: exploratory study of a procedure for persons with severe mental retardation. European Journal of Psychology of Education,19(4), 349-364.

Schraw, G. (2001). Promoting general metacognitive awareness.In H. Hartman (Ed.), Metacognition in learning and instruction. Theory, research and practice (pp. 3–16). Dordrecht: Kluwer.

Schraw, G., &Dennison, R. S. (1994). Assessingmetacognitiveawareness. Contemporary Educational Psychology, 17, 460–475.

Schraw, G., Dunkle, M.E., Bendixen, L.D. & Roedel, T.D. (1995). Does a general monitoring skill exist? Journal of Educational Psychology,87: 433–444.

Schraw, G., & Moshman, D. (1995).Metacognitive theories. Educational Psychology Review,7(4), 351-371.

Snowman, J. & McCown, R. (2012). Psychology Applied to Teaching. (13th edition) Wadsworth Cengage Learning.

Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. Contemporary Educational Psychology,27(1), 51-79.

Sperling, R.A., Richmond, A.S., Ramsay, C. M., & Klapp, M. (2012). The measurement and predictive ability of metacognition in middle school learners. Journal of Educational Research,105(1), 1-7.

Stephens, M. (1998). AAMT discussion paper on assessment and reporting in school mathematics, The Australian Mathematics Teacher, 44(1), 16a-16c

Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. Journal of Educational Psychology,82(2), 306-314.

van der Stel,M., & Veenman,M. V. J. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. Learning and Individual Differences, 18(1), 128–134.

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method design? In C. Artelt, & B. Moschner (Eds), Lernstrategien und Metacognition : Implikationen fu ̈r Forschung und Praxis (pp.77-99). Derlin: Waxmann

Veenman, M. V. J., van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. Metacognition and Learning, 1(1), 3–14.

Wang, C. Y., (2015). Exploring General Versus Task-Specific Assessments of Metacognition in University Chemistry Students: A Multitrait–Multimethod Analysis. Res Sci Educ, 45, 555-579.

Watt, H. M. G. (2005). Attitudes to the use of alternative assessment methods in mathematics: a study with secondary mathematics teachers in Sydney, Australia. Educational Studies in Mathematics,58(1), 21-44.

Weinstein, C. E., Palmer, D. R.,&Schulte, A. C. (1987). Learning and study strategies inventory. Clearwater, FL: H&H Publishers.

Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. Journal of Educational Psychology, 89(3), 397-410