OER FINAL REPORT SERIES

The OER Final Report series includes final reports from funds managed by Office of Education Research, National Institute of Education, Nanyang Technological University.

Reports are submitted as part of the funding review process and intended for the funding agency, local schools and educators, teacher educators, policymakers, and education scholars. They do not take the place of scholarly, peer-reviewed articles but report on the background, procedures, and major findings of the project.*

This study was funded by Singapore Ministry of Education (MOE) under the Education Research Funding Programme (OER 01/16 CBL) and administered by National Institute of Education (NIE), Nanyang Technological University, Singapore. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Singapore MOE and NIE.

*In some cases reports show coloured font or highlights. These are an artifact of the review process and not intended to have any special weight or meaning within the report itself.

Version: 17 Jul 2020, NIE OER

EDUCATION RESEARCH FUNDING PROGRAMME

PROJECT CLOSURE REPORT

Justification in Mathematics (JiM)

By

Assistant Professor Chua Boon Liang

National Institute of Education
Singapore

EXECUTIVE SUMMARY

INTRODUCTION/BACKGROUND

This JiM project set out to research the current state of understanding of mathematical justification by both Singapore secondary school students and Mathematics teachers across the different content strands in the Singapore secondary Mathematics curriculum. This timely and relevant project was started to complement the current goal of the Ministry of Education (MOE) on fostering reasoning in the learning of Mathematics in schools. While MOE focuses on the reasoning process, our study examined the outcomes of this reasoning process.

For many years, mathematical reasoning and communication have been two key process skills in the Singapore Mathematics Framework under the *Processes* component. Thus, mathematical reasoning and justification are essential components of any mathematical activity that students should be familiar with. Yet from our personal communications with secondary school mathematics teachers between 2013 and 2016 and evidence from GCE O and N Level examiners' reports, we obtained the impression that many students, as well as mathematics teachers, struggled to navigate mathematical justification successfully. Mathematics teachers seemed to point students' difficulties in justification to several student-related factors. Other than those factors, we, however, speculated that there might be other potential obstacles, which contribute to students' poor performance in justification tasks. Therefore, this JiM project was undertaken to explore the quality of written justifications constructed by secondary school students and their mathematics teachers, with the aim of proposing a professional development (PD) programme that would allow the teachers to better engage their students in mathematical reasoning and justification more effectively, thereby improving the quality of written justifications provided by their students.

STATEMENT OF PROBLEMS

Given MOE's growing emphasis on equipping students with 21st century competencies, greater demands have been placed on them to reason, explain and justify in the learning of mathematics. While these skills are indeed crucial in scaffolding students' thinking and reasoning, our personal communications with secondary school mathematics teachers showed that many perceived justification tasks in GCE O level and N level examinations as higher-order thinking questions intended for the more able students. Further, students also tend to avoid doing such questions in tests and examinations. It is thus for these reasons that make mathematical justification a worthwhile area to study.

PURPOSE OF STUDY

To unravel the understanding of mathematical justification by both secondary school students and mathematics teachers, this JiM project took on an empirical investigation with the following aims:

(i) to *examine the quality of written justifications* produced by Express students, Normal (Academic) students and mathematics teachers through administering a paper-and-pencil test,
(ii) to *investigate the participants' views of an acceptable justification* through a survey,

The research questions were:
RQ1 What is Singapore secondary school students' understanding of justification?

RQ2 What is Singapore mathematics teachers' understanding of justification?

RQ3 Does Singapore mathematics teachers' understanding of justification predict Singapore Secondary school students' understanding of justification?

PARTICIPANTS

The JiM project took place between July 2016 to Dec 2018. Data for the main study were gathered between February 2017 to October 2017. The main study involved 322 Secondary 4 students (165 Express, 157 Normal (Academic)) and 45 mathematics teachers in five secondary schools. The participating schools were recruited through voluntary application in response to our call for participation. Their mean PSLE aggregate scores range between 199 and 221 for Express and between 161 and 182 for Normal (Academic). The teachers were chosen by the respective schools following these criteria: (a) teaching/have taught different grade levels, (b) the number of years of teaching secondary mathematics, and (c) gender.

METHODOLOGY / DESIGN

The JiM project adopted a survey design to collect data using two instruments: the JiM Test and the JiM Survey.

(a) ***The JiM Test*** is a paper-and-pencil test administered to examine the ways the participating Secondary 4 students and mathematics teachers expressed justifications and perceived the questions' level of difficulty. There were nine questions comprising 12 test items. The test items covered all three content strands in the Singapore Mathematics syllabus: Number and Algebra, Geometry and Measurement, and Statistics and Probability, and involved three types of justification tasks, namely *validation*, *decision-making* and *inference.* The student's version and the teacher's version of the test contained the same set of 12 test items but differed only in instructions. The teachers were asked not only to write down the justifications in the way they hoped their students would produce in order to deserve the highest score but also to indicate their average graduating students' perception of the questions' level of difficulty. The test duration was an hour for both the participating students and the teachers. Calculator use was allowed.

(b) ***The JiM Survey*** is a questionnaire designed to explore what the participating Secondary 4 students and mathematics teachers perceived as an acceptable justification. The questionnaire items were related to the 12 test items in the JiM Test. For each test item in the JiM Test, four possible student justifications picked from the JiM Test completed earlier were provided. The 12 questionnaire items was divided into three sets for the students' version with four items in each while the teacher's version was divided into two sets with six items in each. Due to the fact that the participants needed time to read each of the four possible student justifications in each questionnaire item, each participating student was asked to complete just only a set of the questionnaire. The same practice applies to the participating mathematics teachers. The participating students had to pick the one(s) that they believed would be awarded the highest mark by their mathematics teacher. On the other hand, the participating teachers would select the one(s) that they believed they would award the highest mark. The students were given 50 minutes for each set while the teachers were given an hour to complete each set of questionnaire.

Follow-up interviews After all the five schools had completed the JiM Test, at least 10% of the Express students (n = 19), Normal (Academic) students (n = 19) and mathematics teachers (n = 24) were interviewed to gather additional information about their reasoning not illuminated in their justifications. Both the students and mathematics teachers that produced exemplary or partially correct or incorrect justification were mostly selected for the interviews.

Each student interview took about 30 minutes while the teacher interview took about 30 to 45 minutes. The interviews were audiotaped for data analysis.

The research data set comprised: student responses to JiM Test items and survey items, teacher responses to JiM Test items and survey items, and interviews of students and teachers. A four-point coding-and-marking scheme was specially developed to code the responses to the JiM Test items. This scheme provided an assessment of the quality of justifications constructed by the participating students and mathematics teachers. A score of 3 denoted either an exemplary response or competent response, a score of 2 was for a response with minor flaws, a score of 1 for a response with major flaws, and a score of 0 for a wrong response, an irrelevant response or no response. The data were then analysed using a mixed-methods approach that involved both quantitative (frequencies, percentages and facility index) and qualitative (thematic) techniques.

FINDINGS / RESULTS

1. There is evidence that participating students in both Express and Normal (Academic) courses found it challenging to handle justification tasks. The struggle with mathematical justification was particularly noticeable in the Normal (Academic) course.

   For the participating Express students,
   (a) only two test items were answered well, with over half of them producing exemplary or competent responses;
   (b) an equal number of the remaining ten test items had between 10% and 30% of them or fewer than a tenth of them producing exemplary or competent responses;
   (c) One of the test items (Question 1) is similar to a popular GCE O Level mathematics examination question. Yet many were not taught the correct method of answering it.

   For the participating Normal (Academic) students,
   (a) the majority of test items, except for three, were rather poorly answered with nearly less than 5% of the students producing exemplary or competent responses;
   (b) those three test items had between 10% and 30% of the students producing exemplary or competent responses.

2. A moderate number of participating mathematics teachers demonstrated a reasonable competency in answering justification tasks. Some test items were answered quite well, but there were also some more demanding ones that challenged them.

   (a) Only three test items had over half of them producing exemplary or competent responses. Another five test items with between 30% and 50% of them producing exemplary or competent responses. Of the remaining four test items, three had between 10% and 30% of them producing exemplary or competent responses and only one had less than a tenth;
   (b) Many of them did not seem to know how much to write to construct an exemplary justification. They tended to gauge the amount to write in the justification based on the mark allocated to the justification task in the test or examination.
   (c) A couple of misconceptions were uncovered. Mathematics teachers confused between Pythagoras Theorem and its converse in Question 1, and consequently they taught their students the wrong way of answering Question 1. Many were also unaware that using a counter-example in Question 2 is enough to disprove a mathematical claim.

3. ***Perceived Level of Difficulty of Test Items***
   (a) In general, the participating Express students did not perceive the majority of the test items as difficult, except for just two test items.

(b) Many participating Normal (Academic) students somewhat did not perceive the test items as easy. The majority of them indicated that nearly half of the test items were difficult.
(c) Most of the participating teachers perceived nearly all the test items easy, except for only one test item which many indicated as tough. They also did not perceive many of the test items to be easy for the average graduating students.
(d) When compared with the facility index which measures the test item's level of difficulty, the perception of the participating students and mathematics teachers was reasonable.

4. ***Justification Tasks by Content Strands***
(a) Probability and Statistics justification tasks were most well-answered by the participating Express students. This is followed by Geometry and Measurement justification tasks and then Number and Algebra justification tasks.
(b) The participating Normal (Academic) students fared equally well in Probability and Statistics justification tasks as well as Geometry and Measurement justification tasks. Tasks involving Number and Algebra were found to be the most challenging.
(c) The participating mathematics teachers generally fared equally well in all three content strands.

5. Justification Tasks by Task Nature
The *Validation* and *Inference* types of justification tasks were answered well by the participating students in both courses and mathematics teachers. But the *Decision-Making* type proved to be tough for all participants, with many not stating clearly a conclusion.

6. Best-mark Justification in Survey
(a) A sizeable number of participating students in both courses and mathematics teachers were able to identify the exemplary response in each questionnaire item.
(b) There was also a moderate number of participating mathematics teachers who would award the highest mark to a flawed justification in eight of the questionnaire items. They failed to spot the major flaw in the justification.

CONTRIBUTIONS

Theory. This JiM project sheds light on the construction of justification by secondary school students and mathematics teachers for the different types of justification tasks across the three content strands. Through analysing the numerous justifications produced by the participating students and mathematics teachers, we believe we have validated the critical features that must be present in an exemplary or competent response for each type of justification task. Mathematics teachers can use these critical features to guide their students to construct exemplary justifications. Another significant contribution is the development of a 4-point coding scheme to classify the various written justifications that were produced. This coding scheme offers a quick means to help mathematics teachers sort and group the student justifications for classroom discussion.

Capacity building of teachers. The participating mathematics teachers benefited not only from their collaboration and interactions with the researchers, but also from the seminars and workshops conducted to share the findings from the project. Other mathematics teachers also benefitted when the team was invited to share the JiM findings at events such as the Math HOD Meetings at MOE and AST's Math Day. Furthermore, the justifications produced by the participating students and mathematics teachers revealed some misconceptions by both students and teachers. These misconceptions were highlighted at the seminars, workshops and invited talks. It is worthwhile to highlight that the JiM project has motivated a Master Teacher to start an NLC group on Mathematical Justification, in which the PI of this JiM project is assisting and advising. Furthermore, the JiM project had drawn interest from a researcher at UCL Institute of Education in the UK to start a small-scale research project between the two countries. In this small-scale project, the PI of the JiM project worked with a local secondary school to mentor two mathematics teachers to develop a series of lessons infusing mathematics justification. The two mathematics teachers were also asked to present the findings of this small-scale project at the Mathematics Teacher Conference in June 2019.

Following the data analysis, we think a reasonable PD programme on fostering mathematical justification in the classroom should comprise the following three phases:

Phase 1: Introduce mathematics teachers to the different types of justification tasks and expose them to the expected demands of each type of task. Get the teachers to attempt the different types of justification tasks, then discuss the responses produced, highlighting the strengths and in particular any weakness so that they can improve on it.
Phase 2: Assign one to two justification tasks for mathematics teachers to implement with their students. Get the teachers to attempt to classify the student justifications, and identify the problems in each category. Invite them to share their classification scheme.
Phase 3: Explore with the mathematics teachers how to promote reasoning and justification in the classroom, in particular, how to move from the oral discussion in class to produce an acceptable written justification. This phase of classroom practice seeks to improve teaching effectiveness and student achievement. (This is a potential area for research for the next stage of the JiM project.)

Practice. Through this study, the participating students and mathematics teachers were exposed to the different types of justification tasks and be acquainted with their demands, which could serve as a guide for them in answering them and, for the teachers, setting questions for their students that would promote mathematical reasoning and justification. Such questions could also be used to strengthen students' foundational understanding in the different content strands. Moreover, the seminars, workshops and invited talks offered the mathematics teachers some ideas and strategies to tackle with justification tasks.

CONCLUSION

The JiM study endeavours to work collaboratively with schools in improving the teaching and learning of mathematical reasoning and justification. Through its efforts, we believe we have reached out to some mathematics teachers to engage and expose their students more frequently to justification tasks, so that with time and constant practice, the students will be able to produce better quality written justifications. To certain extent, we had better equipped mathematics teachers involved in the JiM project in addressing their misconceptions and errors and honing their reasoning, communication and writing skills. This would in turn greatly aid them in producing quality written justifications and will make the learning of mathematical justification a more enriching and less intimidating process for them and for their students.

ACKNOWLEDGEMENTS

This JiM project was funded by the Education Research Funding Programme, National Institute of Education (NIE), Nanyang Technological University, Singapore, project no. OER 01/16 CBL. The views expressed in this report are the author's and do not necessarily represent the views of NIE.

KEYWORDS

Mathematical Justification, Mathematical Reasoning, Secondary Mathematics, Students' Perception, Teachers' Perception

Justification in Mathematics (JiM)

Assistant Professor Chua Boon Liang

National Institute of Education

INTRODUCTION/BACKGROUND

Mathematical reasoning and communication have been two key processes under the *Processes* component in the Singapore Mathematics Framework for many years. While mathematical reasoning refers to the ability of analysing mathematical situations and constructing logical arguments (Ministry of Education, Singapore, 2012), mathematical justification, which is a form of communication, involves articulating the mathematical ideas or arguments in a precise, concise and logical manner using appropriate mathematical language. However, the extent to which students are exposed to justification tasks varies, as mathematics teachers seem to perceive such justification tasks to be higher-order thinking questions meant for the more able students. Moreover, students tend to avoid such tasks in the GCE N and O level examinations owing to difficulties encountered when answering such tasks.

In a bid to unravel the understanding of mathematical justification by both secondary school students and mathematics teachers and understand the challenges faced by students when dealing with justification tasks, the JiM project was thus conceptualised. It sought to conduct an empirical investigation into how Secondary Four students and mathematics teachers constructed written justifications and also find out from the teachers on what exactly they expected from the students for different types of justification tasks. This project was timely as the students were preparing for their upcoming GCE N and O level examinations, where questions requiring them to justify were common. Through follow-up interviews with selected participating students and mathematics teachers as well as sharing sessions with the mathematics teachers that allowed the JiM research team to shed light on not only the misconceptions held by both the participating students and even teachers, but also the mistakes made in their written justifications. This enabled the mathematics teachers to convey the feedback to their students, so that they could rectify their mistakes and not repeat them in the examinations.

In the second part of the study, the perceptions of the participating students and mathematics teachers on what constitutes an “acceptable justification” were explored. It also aimed to investigate whether the participating students and mathematics teachers were able to distinguish an exemplary justification from another one that had minor/major flaws. The findings from the two phases were then used in proposing a professional development (PD) programme to support mathematics teachers as they engage and help students to develop competencies in mathematical reasoning and justification, which in turn would improve the quality of students’ written justifications with sufficient time and practice.

STATEMENT OF PROBLEMS (I.E., JUSTIFICATION FOR THE STUDY)

Given MOE's growing emphasis on equipping students with 21st century competencies, greater demands have been placed on them to reason, explain and justify in the learning of mathematics. These skills are particularly crucial in probing students' understanding as they come up with a series of arguments, as well as validating and explaining their computations leading to the solution. Furthermore, those skills also enable students to integrate their new knowledge with existing knowledge, as they draw links to other mathematical ideas. They also make students' thinking visible so that misconceptions could be spotted and addressed promptly by mathematics teachers. In the process of reasoning and justification, it is also crucial for students to possess communication skills, which involves using the language of mathematics to express mathematical ideas and arguments in a clear and concise manner.

While these skills are indeed crucial in scaffolding students' thinking and reasoning, our personal communications with secondary school mathematics teachers showed that many perceived justification tasks in GCE O level and N level examinations as higher-order thinking questions intended for the more able students. Moreover, students also tend to give such questions a miss in these national examinations. Previous studies such as that of Kuchemann and Hoyles (2003), Stylianides (2015) and Chua (2016) have found that despite students having the right idea, they often experience difficulty in expressing their justifications in written form, clearly and concisely. Moreover, Healy and Hoyles (2000) study showed that high-ability students aged 14-15 were better at choosing correct mathematical justifications than constructing them, as they were not acquainted with different methods of justifying. With the Singapore mathematics curriculum increasingly steering towards mathematical reasoning and justification, mathematics teachers' "negative" perceptions and the challenges faced by students when tackling justification tasks are indeed an area of concern. Therefore, in a bid to effectively address these challenges and improve the quality of written justifications produced by students, the JiM project emerged.

Seeking to investigate the current state of written justifications produced by Secondary Four Singapore students and mathematics teachers, this JiM project served as a platform to

(a) spot the difficulties encountered by both secondary school students and mathematics teachers when answering justification tasks; and
(b) unearth what mathematics teachers and students perceive as "acceptable justifications";

The findings would help the research team to come up with their PD programme for mathematics teachers to enable them to engage and help students to produce good quality written justifications.

PURPOSE OF STUDY (INCLUDING RESEARCH QUESTIONS AND/OR OBJECTIVES)

To unravel the understanding of mathematical justification by both secondary school students and mathematics teachers, this JiM project took on an empirical investigation with the following aims:

(i) to *examine the quality of written justifications* produced by Express students, Normal (Academic) students and mathematics teachers through administering a paper-and-pencil test,
(ii) to *investigate the participants views' of an acceptable justification* through a survey,

The research questions were:

RQ1 What is Singapore secondary school students' understanding of justification?

RQ2 What is Singapore mathematics teachers' understanding of justification?

RQ3 Does Singapore mathematics teachers' understanding of justification predict Singapore Secondary school students' understanding of justification?

PARTICIPANTS

The JiM project took place between July 2016 to Dec 2018. Data for the pilot study and the main study were gathered between January 2017 and February 2017, and between February 2017 to October 2017 respectively.

The pilot study involved 30 Secondary Four Express students, 27 Normal (Academic) students and 10 mathematics teachers in one secondary school. The main study involved 165 Express students, 157 Normal (Academic) students and 45 teachers in five secondary schools. The breakdown of the number of Secondary Four students and mathematics teachers are shown in Table 1.

Table 1. *Number of Secondary Four students by school and course of study*

School	Express	Normal (Academic)	Teachers
School A (Pilot)	30	27	10
School B	39	40	6
School C	32	38	10
School D	33	31	11
School E	37	34	10
School F	24	14	8
Total	195	184	55
	379 (178 boys, 201 girls)		55 (29 males, 26 females)

The six participating schools were recruited through voluntary application in response to our call for participation. For the five schools in the main study, their mean PSLE aggregate scores range between 199 and 221 for Express and between 161 and 182 for Normal (Academic).

The participating teachers in both the pilot study and the main study were chosen by the respective schools following these criteria: (a) teaching/have taught different grade levels, (b) the number of years of teaching secondary mathematics, and (c) gender. The breakdown of the number of mathematics teachers by the years of teaching experience is shown in Table 2.

Table 2. *Number of mathematics teachers by years of teaching experience*

Years of Teaching Experience Gender	t < 2	2 ≤ t < 5	5 ≤ t < 10	10 ≤ t < 15	t ≥ 15	Total
Male	3	7	6	7	6	29
Female	1	2	6	9	8	26
Total	4	9	12	16	14	55

23.6% of the participating mathematics teachers had been teaching for less than 5 years, another 50.9% for at least five to less than 15 years, and the remaining 25.5% for 15 years or more.

METHODOLOGY/DESIGN

The JiM project used a survey design to collect data using two instruments: (a) the JiM Test and (b) the JiM Survey.

(a) The JiM Test

This JiM test is a paper-and-pencil test administered to examine the ways the participating Secondary 4 students and mathematics teachers expressed justifications and perceived the questions' level of difficulty. There were nine questions comprising 12 test items. The test items covered all three content strands in the Singapore Mathematics syllabus: Number and Algebra, Geometry and Measurement, and Statistics and Probability, and involved three types of justification tasks, namely *validation*, *decision-making* and *inference.* There were four test items for each content strand. Of the 12 test items, three were *validation* type, five were *decision-making* and the remaining four were *inference*.

The student's version and the teacher's version of the test contained the same set of 12 test items but differed only in instructions. The teachers were asked not only to write down the justifications in the way they hoped their students would produce in order to deserve the highest score but also to indicate their average graduating students' perception of the questions' level of difficulty. The test duration for the students was an hour while for the teachers, it was 45 minutes. Calculator use was allowed. Table 3 below lists the 12 test items by content strands and nature of the task.

Table 3. Justification tasks *by content strands and nature of task*

Task	Name of Task	Content Strand	Task Nature
Question 1	Let's Join The Dots	Geometry and Measurement	Validation
Question 2	Balloon	Number and Algebra	Decision-making
Question 3a	Let's Go Jogging	Number and Algebra	Inference
Question 3b	Let's Go Jogging	Number and Algebra	Inference
Question 4a	Children's Day Gifts	Probability and Statistics	Validation
Question 4b	Children's Day Gifts	Probability and Statistics	Decision-making
Question 5a	Rectangles	Geometry and Measurement	Decision-making
Question 5b	Rectangles	Geometry and Measurement	Inference
Question 6	Number on Dice	Probability and Statistics	Inference
Question 7	Income	Probability and Statistics	Decision-making
Question 8	Curve3	Number and Algebra	Validation
Question 9	Rooftop Garden	Geometry and Measurement	Decision-making

(b) The JiM Survey

The JiM survey a questionnaire designed to explore what the participating Secondary 4 students and mathematics teachers perceived as an acceptable justification. The questionnaire items were related to the 12 test items in the JiM Test. For each test item in the JiM Test, four possible student justifications picked from the JiM Test completed earlier were provided. The 12 questionnaire items was divided into three sets for the students' version with four items in each while the teacher's version was divided into two sets with six items in each. Due to the fact that the participants needed time to read each of the four possible student justifications in each questionnaire item, each participating student was asked to complete just only a set of the questionnaire. The same practice applies to the participating mathematics teachers. The participating students had to pick the one(s) that they believed would be awarded the highest mark by their mathematics teacher. On the other hand, the participating teachers would select the one(s) that they believed they would award the highest mark. The students were given 50 minutes for each set while the teachers were given an hour to complete each set of questionnaire.

(c) Procedures

For the main study, the research team intended to administer the JiM test on the same day for all the participating students and mathematics teachers. However, owing to the schools' tight schedules, the test was administered on separate days for most of the schools. The research team particularly advised the mathematics teachers to discourage their students from discussing the answers with their peers, if they had not taken the test at the same time.

This was to ensure that the responses produced would be valid for the study. The participating students were seated in the order of their registration number in the classlist. Both students and teachers were given 60 minutes to complete all the 12 test items in the JiM Test. However, in the event that they were unable to complete the test within the stipulated time, extra time was given. Prior to participating in this study, all the participating students had learnt all the mathematical concepts tested in the test. After the test, the research team collected all the scripts and then started looking through the responses to select students and teachers for follow-up interviews.

For the interviews, the research team ensured that they selected an equal (or approximately equal) number of Express and Normal (Academic) students from each school to ensure that both courses of study were equally represented. Among the mathematics teacher interviewees, it was ensured that the Head of Department or Subject Head of Mathematics who took the test, was selected for the follow-up interview. The reason being that they played an important role in shaping the manner mathematics should be taught in their schools. We hope to understand how they would assess particular responses and their receptiveness towards the participating students and mathematics teachers adopting alternative methods of solving a particular question (even though such a method may not have been covered in the syllabus).

In total, 19 Express students, 19 Normal (Academic) students and 24 teachers were selected for the follow-up interviews, in order to gather additional information about their reasoning not illuminated in their justifications. Participating students and mathematics teachers who produced exemplary or partially correct or incorrect justification were mostly selected for the follow-up interviews. Depending on the school schedule, each student interview took about 30 minutes while the teacher interview took about 30 to 45 minutes. The interviews were audiotaped for data analysis.

While the follow-up interviews were being carried out, the written responses produced in the JiM Test were coded and analysed using a specially developed coding-and-marking scheme that underwent a number of cycles of revision. The justifications constructed by the participating students and mathematics teachers for each test item in the JiM Test were coded and scored using a four-point scale: 0 point to 3 points. This coding-and-marking scheme provided an indication of the quality of justifications constructed by the students and mathematics teachers. A score of 3 denoted either an *exemplary* response or *competent* response, a score of 2 for a response with minor flaws, a score of 1 for a response with major flaws, and a score of 0 for a wrong/irrelevant response or no response.

The research team engaged a student assistant who was a Mathematics Major at NIE to code the student responses. One entire class of student responses was randomly selected for the first round of coding. The coding-and-marking scheme then underwent some revision and refinement. The revised scheme was subsequently used to code another different class of student responses. This cycle of refinement stopped after the scheme had somewhat stabilised. Using the stablished version of the coding-and-marking scheme, all the student responses were recoded. Likewise, the mathematics teachers' responses were coded using the stablished scheme. An inter-coder check was performed on all the responses produced by the participating students and mathematics teachers with the very kind assistance of a mathematics PhD candidate at the NIE. All the test items achieved over 90% consistency. After the coding was completed, a frequency count was done for each test item to determine the number of participating students and mathematics teachers who produced a particular level of response.

The second phase of the JiM study was marked by the JiM Survey, which was administered via a questionnaire, to uncover the perceptions of an acceptable justification by the participating students and mathematics teachers. Not all the participating students and mathematics teachers who took the JiM Test partook in the survey due to various reasons. In the main study, 158 Express students, 144 Normal (Academic) students and 46 teachers took the survey. Tables 4 and 5 reports the breakdown of the number of participating students and mathematics teachers taking the different sets of questionnaires. Keeping in mind that unlike the JiM Test, the participating students and mathematics teachers had to answer fewer questions (students were to answer four questions while the teachers were to answer six questions), students were given 45 minutes to complete the questionnaire while the teachers were still given 60 minutes as they had more questions to answer compared to their students. However, it was not surprising that most of the teachers managed to finish the questionnaire in less than 60 minutes. After the survey, the results were collated and a frequency count was done to determine the number of students and teachers who chose a particular student response to be awarded the highest mark.

Table 4. *Number of participating students surveyed by questionnaire set*

Questionnaire Set	Express	Normal (Academic)
Set A	53	54
Set B	53	46
Set C	52	44
Total	158	144

Table 5. *Number of participating mathematics teachers surveyed by questionnaire set*

Questionnaire Set	Teachers
Set P	24
Set Q	22
Total	46

(d) Data Analysis

The research data set comprised: student responses to JiM Test items and survey items, teacher responses to JiM Test items and survey items, and interviews of students and teachers. A four-point coding-and-marking scheme was specially developed to code the responses to the JiM Test items. This scheme provided an indication of the quality of justifications constructed by the participating students and mathematics teachers. Figure 1 below provides an example of the coding-and-marking scheme for the *Balloon* task in Question 2. A score of 3 denoted either an exemplary response or competent response, a score of 2 was for a response with minor flaws, a score of 1 for a response with major flaws, and a score of 0 for a wrong response, an irrelevant response or no response.

The data were then analysed using a mixed-methods approach that involved both quantitative (frequencies, percentages and facility index) and qualitative (thematic) techniques. The facility index was used to calculate the average score obtained by the participating Express, Normal (Academic) and mathematics teachers for a particular test item. The formula is given as follows:

$$Facility\ Index = \frac{Total\ marks\ obtained\ by\ all\ students\ (teachers)}{Total\ number\ of\ students\ (teachers)} \times 100\%$$

Table 6 shows the facility index and the test item's level of difficulty.

Table 6. *Facility index and level of difficulty*

Facility Index (%)	Level of Difficulty
0 ≤ x ≤ 20	Very Difficult
20 < x ≤ 40	Difficult
40 < x ≤ 60	Moderate
60 < x ≤ 80	Easy
80 < x ≤ 100	Very Easy

Question 2: ***Balloon*** (Number and Algebra, Decision-making)

Jeya fills a balloon with some air and at the same time measures the pressure, *P*, and volume of air, *V*, inside the balloon.
He observes that for constant room temperature, the product of the pressure and the volume of air in the balloon remains constant. That is, $PV = k$, where *k* is a constant.

Jeya then decreases the volume of air inside the balloon by 10% while maintaining the same temperature.

Explain whether the pressure of air inside the balloon now increases by 10%.

Level		**Descriptor**	**Examples of responses**
Level 3	Exemplary	• Using letters to explore relationship between initial and new P Show $P_{new}V_{new} = P_{original}V_{original}$ $P_{new}(\frac{90}{100}V_{original}) = P_{original}\ V_{original}$ Hence, $P_{new} = \frac{100}{90}P_{original}$ • A decision is made regarding the claim So, P does not increase by 10% but by $11\frac{1}{9}\%$ (or P increases by $11\frac{1}{9}\%$ and not 10%).	Acceptable workings: 1. $V_{new} = \frac{9}{10}V_{original}$ If $P_{new} = \frac{11}{10}P_{original}$, then $P_{new}V_{new} = \frac{11}{10}P_{original} \times \frac{9}{10}V_{original}$ $= \frac{99}{100}\ P_{original}V_{original}$ which is not constant (should obtain $P_{new}V_{new} = P_{original}V_{original}$ if the claim is true) 2. If P increases by 10%, then $P_{new} = \frac{110}{100}P_{original}$. But $P_{new} \neq \frac{110}{100}P_{original}$.
	Competent	• Use a specific case to test the claim	Acceptable workings: P = 100, V = 10, k = 1000

		Use a particular set of numerical values for P and V to work out the percentage change in P and show that it did not increase by 10% E.g, Let P = 100, V = 10 So k = 1000 When V = 9 (that is decreased by 10%), P = $\frac{1000}{9} = 111\frac{1}{9}$ • A decision is made regarding the claim So, P does not increase by 10% but by $11\frac{1}{9}\%$ (or P increases by $11\frac{1}{9}\%$ and not 10%).	$V_{new} = 9$ If P increases by 10%, $P_{new} = 110$. $P_{new} \times V_{new} = 100 \times 9 = 990$ which is not equal to k = 1000 or $P_{new} = \frac{1000}{9}$ which is not 110 ∴ P does not increase by 10% Acceptable: If P increases by 10%, then $P_{new} = \frac{110}{100}P$. But their $P_{new} \neq P_{new}$
Level 2	Satisfactory with minor flaws	• Similar to Level 3, except for (i) not making decision about the claim despite correct working (e.g., $P_{new} = \frac{10}{9}P_{original}$ or equiv seen), or no evidence of actual % increase (ii) giving vague or contradictory conclusion (iii) error in presentation (e.g., % sign missing) or slip in computation (e.g., SYH, Exp, #27) Or	

		• Similar to code 32, except for rounding down actual % / value of P to nearest multiples of 10 (e.g., from $111\frac{1}{9}$ to 110, or $11\frac{1}{9}$% to 10%) and then concluding that P increases by 10%.	
Level 1	Below satisfactory with major flaws	• Somewhat indicate that the new value of V is $\frac{9}{10}$ the original value of V. • Never distinguish new and original values of P and V clearly, resulting in P = $\frac{k}{0.9V}$ (or $\frac{k}{\frac{9}{10}V}$, $\frac{P}{Pnew} = \frac{0.9V}{V}$ seen) and getting stuck here. • May or may not give a clear decision about P. The decision, if given, may or may not be correct.	
Level 0	Irrelevant, incorrect blank	e.g., Indicate that when V decreases by 10%, P has to increase 10% to keep k constant.	
	Misread of givens	e.g., V is misread as having increased by 10%, working is otherwise correct	

Figure 1. *Coding-and-marking scheme for Balloon Task*

FINDINGS / RESULTS

1. Producing Exemplary or Competent Justifications

The data analysis shows compelling evidence that the participating students in both Express and Normal (Academic) courses found it challenging to handle justification tasks. The struggle with mathematical justification was particularly noticeable in the Normal (Academic) course.

For the participating Express students,

(a) over half of them produced exemplary or competent responses in only two test items (Questions 4(a) and 5(b));
(b) less than a tenth of them produced exemplary or competent responses in five test items (Questions 2, 3(b), 4(b), 7 and 9);
(c) between 10% and 30% of them produced exemplary or competent responses in the remaining five test items (Questions 1, 3(a), 5(a), 6, and 8).

For the participating Normal (Academic) students,

(a) between 10% and 30% of them produced exemplary or competent responses in just three test items (Questions 4(a), 5(b) and 6);
(b) nearly less than 5% of them produced exemplary or competent responses in the remaining nine test items (Questions 1, 2, 3(a), 3(b), 4(b), 5(a), 7, 8 and 9).

Two questions are worth highlighting here. A striking finding from the analysis of the responses produced by Express students revealed a gross error in Question 1, which asked the participants to explain why a triangle is a right-angled triangle. This is a popular question in the GCE O Level examination. A vast majority of the Express students did not test whether or not the sum of the squares of each of the two shorter sides is equal to the square of the longest side (which is the Converse of the Pythagoras Theorem). Instead, they started by assuming the triangle is right-angled, then verified that the equality was true. This way of justifying that the triangle is right-angled is wrong. Another surprising finding is observed in Question 2 (see Figure 1 above). Many students struggled with the algebraic manipulation when the mathematical claim could be disproved easily by using a numerical counter-example.

A moderate number of participating mathematics teachers demonstrated a reasonable competency in answering justification tasks. Most of the test items were answered well by many teachers, but there were also some more demanding ones that challenged them.

(a) Over half of them produced exemplary or competent responses in three test items (Questions 2, 5(b) and 6).
(b) Between 30% and 50% of them produced exemplary or competent responses in five test items (Questions 1, 3(b), 4(a), 5(a) and 8).
(c) Of the remaining four test items, three had between 10% and 30% of them producing exemplary or competent responses (Questions 3(a), 7 and 9) and only one had less than a tenth (Question 4(b)).
(d) An interesting finding to emerge from the teacher interviews is that a number of mathematics teachers mentioned about not knowing what makes a complete justification. They tend to gauge how much to write to construct an exemplary or competent justification based on the mark allocated to the justification task in the test or examination (e.g., provide just a point if the justification task is worth only 1 mark). However, this way of determining what an exemplary or competent justification should

entail may not be always accurate and sensible. An exemplary or competent justification is hard to be measured by the mark allocated to the justification task.

(e) Another noteworthy finding is that some mathematics teachers were not clear about certain mathematical ideas. For instance, a sizeable number of mathematics teachers were unable to distinguish a theorem and its converse (e.g., Pythagoras Theorem and its converse in Question 1). As a result, they did not teach their students the correct way of answering Question 1. Many were also not aware that using a counter-example in Question 2 is enough to disprove a mathematical claim. For those who were aware, they discouraged students from using such a method for fear that the students would wrongly apply the method to prove a mathematical claim.

In general, the following three test items appear to be top-scoring justification tasks for both the participating students and mathematics teachers: Questions 4(a), 5(b) and 6. On the other hand, they seem to find these three test items challenging: Questions 4(b), 7 and 9.

2. Perceived Level of Difficulty of Test Items

(a) In general, the participating Express students did not perceive the majority of the test items as difficult, except for just two test items (Questions 2 and 9).

(b) Many participating Normal (Academic) students somewhat did not perceive the test items as easy. The majority of them indicated that nearly half of the test items were difficult (Questions 1, 2, 7, 8 and 9).

(c) Most of the participating teachers perceived nearly all the test items easy, except for only one test item which many indicated as tough (Question 9). They also did not perceive many of the test items to be easy for the average graduating students (Questions 2, 4(b), 5(a), 5(b), 7, 8 and 9) .

(d) When compared with the facility index, the perception of the participating Express students was close. Besides the two test items that they indicated as not easy, their performance in the JiM Test revealed that these three test items were also not easy for them: Questions 1, 4(b) and 7. For Question 1 that tests on the *Converse of Pythagoras' Theorem*, we found out that many of them were taught wrongly by their mathematics teachers, hence the low occurrence of exemplary or competent responses. As for Questions 4(b) and 7 which are both decision-making justification tasks, many did not answer the questions completely by stating a conclusion. As a result, their responses failed to qualify as exemplary or competent. This indicates that the participating students might not be acquainted with the different types of justification tasks and their demands.
The participating Normal (Academic) students and mathematics teachers also had reasonably good and realistic self awareness of the difficulty level of the test items. For the mathematics teachers, they fared poorly in Questions 4(b) and 7 for the same reason as the participating Express students.

3. Justification Tasks by Content Strands

(a) Probability and Statistics justification tasks were most well-answered by the participating Express students. This is followed by Geometry and Measurement justification tasks and then Number and Algebra justification tasks.

(b) The participating Normal (Academic) students fared equally well in Probability and Statistics justification tasks as well as Geometry and Measurement justification tasks. Tasks involving Number and Algebra were found to be the most challenging.

(c) The participating mathematics teachers generally fared equally well in all three content strands.

4. Justification Tasks by Task Nature

The *Validation* and *Inference* types of justification tasks were answered well by the participating students in both courses and mathematics teachers. But the *Decision-Making* type proved to be tough for all participants, with many not stating clearly a conclusion.

5. Best-mark Justification in Survey

(a) A sizeable number of participating students in both courses and mathematics teachers were able to identify the exemplary response in each questionnaire item. This suggests that the participating students were able to identify the correct justifications even though they might not be competent in constructing them since they were not acquainted with the different types of justification tasks.

(b) There was also a moderate number of participating mathematics teachers who would award the highest mark to a flawed justification in eight of the questionnaire items. They failed to spot the major flaw in the justification. This suggests that the teachers would need to be more careful and conscious of flaws when they read and grade their students' justifications.

CONTRIBUTIONS OF STUDY

We think that the findings that emerged from the JiM project can contribute to Singapore Education and the field in the following ways:

Theory This JiM project sheds light on students and teachers' construction of justification for different types of justification tasks across the three content strands. Through analysing the numerous justifications produced by the participating students and mathematics teachers, we believe we have validated the critical features that must be present in a response for each type of justification task. Mathematics teachers can use these critical features to guide their students to construct exemplary justifications. Another significant contribution is the development of a 4-point coding scheme to classify the various justifications that were produced. This coding scheme offers a quick means to help mathematics teachers sort and group the student justifications for classroom discussion.

Capacity building of teachers The participating mathematics teachers benefited not only from their collaboration and interactions with the researchers, but also from the seminars and workshops conducted to share the findings from the project. Other mathematics teachers also benefitted when the team was invited to share the JiM findings at events such as the Math HOD Meetings at MOE and AST's Math Day. Furthermore, the justifications produced by the participating students and mathematics teachers revealed some misconceptions by both students and teachers. These misconceptions were highlighted at the seminars, workshops and invited talks. It is worthwhile to highlight that the JiM project has motivated a Master Teacher to start an NLC group on Mathematical Justification, in which the PI of this JiM project is assisting and advising. Furthermore, the JiM project had drawn interest from a researcher at UCL Institute of Education in the UK to start a small-scale research project between the two countries. In this small-scale project, the PI of the JiM project worked with a local secondary school to mentor two mathematics teachers to develop a series of lessons infusing mathematics justification. The two mathematics teachers were also asked to present the findings of this small-scale project at the Mathematics Teacher Conference in June 2019.

Following the data analysis, we think a PD programme for mathematics teachers is necessary. We propose that a reasonable PD programme on fostering mathematical justification in the classroom should comprise the following three phases:

Phase 1: Introduce mathematics teachers to the different types of justification tasks and expose them to the expected demands of each type of task. Get the teachers to attempt the different types of justification tasks, then discuss the responses produced, highlighting the strengths and in particular any weakness so that they can improve on it.

Phase 2: Assign one to two justification tasks for mathematics teachers to implement with their students. Get the teachers to attempt to classify the student justifications, and identify the problems in each category. Invite them to share their classification scheme.

Phase 3: Explore with the mathematics teachers how to promote reasoning and justification in the classroom, in particular, how to move from the oral discussion in class to produce an acceptable written justification. This phase of classroom practice seeks to improve teaching effectiveness and student achievement. (This is a potential area for research for the next stage of the JiM project.)

Practice Through this study, the participating students and mathematics teachers were exposed to the different types of justification tasks and be acquainted with their demands, which could serve as a guide for them in answering them and, for the teachers, setting questions for their students that would promote mathematical reasoning and justification. Such questions could also be used to strengthen students' foundational understanding in the different content strands. Moreover, the seminars, workshops and invited talks offered the mathematics teachers some ideas and strategies to tackle with justification tasks.

CONCLUSION

The JiM study endeavours to work collaboratively with schools in improving the teaching and learning of mathematical reasoning and justification. Through its efforts, we believe we have reached out to many mathematics teachers to engage and expose their students more frequently to justification tasks, so that with time and constant practice, the students will be able to produce better quality written justifications. To certain extent, we have also better equipped mathematics teachers involved in the JiM project in addressing their misconceptions and errors and honing their reasoning, communication and writing skills. This would in turn greatly aid them in producing quality written justifications and will make the learning of mathematical justification a more enriching and less intimidating process for them and for their students.

ACKNOWLEDGEMENTS

This JiM project was funded by the Education Research Funding Programme, National Institute of Education (NIE), Nanyang Technological University, Singapore, project no. OER 01/16 CBL. The views expressed in this report are the author's and do not necessarily represent the views of NIE.

REFERENCES

Chua, B. L. (2016). Examining mathematics teachers' justification and assessment of students' justifications. In Csíkos, C., Rausch, A., & Szitányi, J. (Eds.) *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education* (pp. 155 - 162). Szeged, Hungary: PME. Heals and Hoyles (2000)

Küchemann, D. & Hoyles, C. (2003). The quality of students' explanations on a non-standard geometry item. In *Proceedings of CERME* (Vol. 3).

Stylianides, A. (2015, February). The role of mode of representation in students' argument constructions. In *CERME 9-Ninth Congress of the European Society for Research in Mathematics Education* (pp. 213-220).