



EXAMINING HIGHER ORDER THINKING IN INDONESIAN LOWER SECONDARY MATHEMATICS CLASSROOMS

Citra Putriarum Tanudjaya, Michiel Doorman

Utrecht University, Utrecht, The Netherlands
Email: c.citraputriarumtanudjaya@students.uu.nl

Abstract

Indonesian students' poor performance in the mathematics test of PISA 2015 prompted the decision by the Ministry of Education of Indonesia to pay more attention to the integration of higher-order thinking (HOT) in the curricula starting in 2018. This new regulation emphasizes the need to have a shared understanding of HOT in mathematics on many levels, such as curriculum, pedagogy, and assessment, and among students, teachers and policy makers. This study aims to examine HOT in Indonesian lower secondary mathematics classrooms by assessing students' ability to demonstrate HOT skills through an open-ended mathematics problem, and by exploring teachers' views of HOT skills through semi-structured interviews. It involved 372 ninth-grade students and six mathematics teachers from six lower secondary schools in Jakarta and Palembang. The findings show that most students could construct the mathematical model but experienced difficulty in transferring knowledge into new contexts, in applying creative thinking, and with information literacy skills. Besides, some of the teachers were familiar with the concept of HOT, but some viewed HOT as skills for talented students, or HOT problems having a high level of difficulty and long storylines. The knowledge of existing teaching strategies, familiarity with HOT problems, and colleague-support are needed to improve the development of HOT skills in the mathematics classroom.

Keywords: Higher-order thinking, Curriculum, Pedagogy, Assessment, Teachers' views, Students' ability

Abstrak

Buruknya performa siswa Indonesia dalam tes matematika PISA 2015 mendorong Menteri Pendidikan dan Kebudayaan Indonesia untuk memberi perhatian lebih atas integrasi keterampilan berpikir tingkat tinggi atau higher-order thinking (HOT) ke dalam kurikulum sejak tahun 2018. Kebijakan baru ini menekankan pentingnya pemahaman akan HOT dalam pelbagai aspek, seperti kurikulum, pedagogi, dan desain penilaian. Penelitian ini bertujuan untuk mengevaluasi penerapan HOT dalam kelas matematika Sekolah Menengah Pertama dengan menilai penguasaan keterampilan HOT siswa melalui masalah matematika terbuka dan dengan mengeksplorasi pemahaman guru akan keterampilan HOT melalui wawancara semi-terstruktur. Partisipan terdiri atas 372 siswa kelas sembilan dan enam guru matematika dari enam Sekolah Menengah Pertama di Jakarta dan Palembang. Hasil penelitian menunjukkan bahwa sebagian besar siswa dapat membuat model matematika dari permasalahan yang diberikan, namun mengalami kesulitan dalam mengembangkan model matematika tersebut sesuai dengan konteks kehidupan sehari-hari. Siswa sulit untuk berpikir kreatif dalam menentukan faktor-faktor yang berpengaruh terhadap model serta memiliki keterbatasan literasi informasi. Selain itu, beberapa guru memahami konsep HOT, namun terdapat kepercayaan bahwa HOT diperuntukkan bagi siswa berbakat, dan bahwa soal HOT memiliki tingkat kesulitan yang tinggi serta alur cerita yang panjang. Pengetahuan mengenai strategi pengajaran, pengenalan akan soal HOT, dan dukungan kolega diperlukan untuk meningkatkan pengembangan keterampilan HOT di kelas matematika.

Kata kunci: Keterampilan berpikir tingkat tinggi, Kurikulum, Pedagogi, Penilaian, Pemahaman guru, Penguasaan siswa

How to Cite: Tanudjaya, C.P., & Doorman, M. (2020). Examining Higher-Order Thinking in Indonesian Lower Secondary Mathematics Classrooms. *Journal on Mathematics Education*, 11(2), 277-300. <http://doi.org/10.22342/jme.11.2.11000.277-300>.

Mathematics reform movements in the 1980s and 1990s put forward a new standard in mathematics education. The National Council of Teachers of Mathematics (1991), the Mathematical Association of America (1991), and the National Research Council (2001) highlighted the importance of having deep

and interconnected understandings of mathematical concepts, procedures, and principles instead of the memorization of formulas and reproducing algorithms. Complete understanding of mathematics requires not only knowledge of mathematical concepts, rules, and their structures, but also the ability to engage in the process of mathematical thinking, which involves reasoning, solving open problems, connecting mathematical ideas, making inferences from data, and communicating mathematics to others (Kitcher, 1984; Lakatos, 1976; National Research Council, 2001; Schoenfeld, 1992). In recent years, some progress has been made in terms of assessment systems in such a way, that they address the complexity of mathematical thinking and problem-solving more thoroughly.

The Organization for Economic Co-operation and Development (OECD), for instance, designs a mathematics literacy test – widely known as the PISA (Programme for International Student Assessment) test – for 15-year-old learners (OECD, 2019b). Besides providing comparable data to improve the participating countries' educational policies and outcomes, the PISA test is intended to describe learners' capacities to reason mathematically and to use mathematical concepts, procedures, facts, and tools to portray, explain and forecast phenomena (OECD, 2018). OECD also highlights the importance for the learners to acquire information literacy skills (Wijaya, 2016). These skills are part of higher-order thinking (HOT) skills, which in turn is used as part of educational reform to mark cognitive activities that are beyond the stage of reproduction (Bloom, 1956). Memorization and recall of information are classified as lower-order thinking, whereas analyzes, syntheses, and evaluation are classified as higher-order thinking skills.

Some countries seem to experience difficulties in preparing their students for the PISA test. For instance, the average mathematics score of Indonesia was positioned at 64th out of 70 participating countries in PISA 2015 (OECD, 2016). This lack of mathematical thinking skills prompted the decision by the Ministry of Education (MOE) of Indonesia to pay more attention to the integration of HOT in the curricula starting in 2018 (Ariyana, Pudjiastuti, Bestary, & Zamroni, 2018). The results of PISA 2018 show that only “around 1% of students in Indonesia can model complex situations mathematically” (OECD, 2019a). For this they lack the skills to select, compare, and evaluate appropriate problem-solving strategies. This recent finding forges the urgency in developing HOT skills among Indonesian students. Although policymakers might be more interested in improving PISA results, it must be noted that placing an emphasis on HOT in education will also improve content area achievement and motivation (Brookhart, 2010), as well as encourage autonomy (Smith & Darvas, 2017). This new regulation emphasizes the need to have shared understanding of HOT in mathematics on many levels, such as the curriculum, the pedagogy, and the assessment design.

Teachers' familiarity with the concept of HOT, nevertheless, is lacking. Many Indonesian teachers argue that there is a lack of information and support from the government to become familiar with HOT (Aini, 2018; Hantoro, 2018; Ahmad et al., 2018; Suryadi, 2018). A small survey by Apino and Retnawati (2017) found that only 20% of high school mathematics teachers in Indonesia applied

teaching practice that aimed to develop HOT skills. A study by Retnawati et al. (2018), although the scope of the study was on the terminology “HOTS” – and not on the definition of the skills, found that mathematics teachers in Indonesia were still unfamiliar with the concept of HOT. Teachers’ understanding of the characteristics of HOT problems, nevertheless, was not identified in this study, which is essential for teachers to be able to provide an appropriate assessment to develop students’ HOT skills.

Meanwhile, Hadi et al. (2018) observed students' difficulties in solving HOT problems and found that one of the constraints the students encountered in solving mathematics is the difficulty in developing a mathematical model. The instrument used in the study, however, was a set of close-ended multiple-choice questions. As closed-ended questions only demand one correct answer, students tend to memorize statements or formulas without deep comprehension of the concept and the course content (Husain, Bais, Hussain, & Samad, 2012). The use of open-ended problems, conversely, helps to encourage students’ creativity and allow various solutions or strategies (Wijaya, 2018).

Hence, before researching on how to improve understanding of HOT skills in mathematics among both students and teachers, this study aims to examine higher-order thinking (HOT) in Indonesian lower secondary mathematics classrooms by answering three research questions. The first research question is to what extent are Indonesian lower secondary students able to demonstrate HOT skills in mathematics based on their abilities to *transfer knowledge into new contexts, apply creative and critical thinking, and problem-solving*? Furthermore, how does Indonesian lower secondary mathematics teachers view HOT skills in terms of *curriculum, pedagogy, and assessment*? Lastly, what are the challenges, needs, and existing supportive factors for Indonesian lower secondary mathematics teachers to develop HOT skills in the mathematics classroom?

This study adopts the definition of HOT as the ability to transfer knowledge into new contexts, apply critical and creative thinking, and problem-solving (Brookhart, 2010), which is in line with the one in the guidebook published by the Indonesian Ministry of Education in 2018, namely *Buku Pegangan Pembelajaran Berorientasi pada Keterampilan Berpikir Tingkat Tinggi* or Learning Handbook Oriented to Higher Order Thinking Skills (Ariyana, Pudjiastuti, Bestary, & Zamroni, 2018). The key findings of the study are aimed to help stakeholders (e.g., Ministry of Education, schools, and mathematics teachers) in identifying challenges and finding solutions to develop students’ HOT skills in the future.

METHOD

Participants

The exploratory study involved 372 ninth-grade students and six mathematics teachers from six lower secondary schools. Three schools are located in Jakarta, and the other three are located in Palembang. Jakarta was chosen because of its variability in educational attainment among both students and teachers. Palembang was selected to also include schools from another area in Indonesia and for

reasons of convenience sampling. Utrecht University has a long-term collaboration with Sriwijaya University (UNSRI) which is located in Palembang. UNSRI has a broad network consisting of lower secondary schools in Palembang, which allows the involvement of secondary teachers. Furthermore, Palembang has a much lower average mathematics score in the national examination which indicates lower educational attainment compared to Jakarta (<https://hasilun.puspendik.kemdikbud.go.id/>). The different characteristics might provide a broader view of the results. The ninth grade was chosen because it is the last grade of junior high school when students take the national examination that partially assesses HOT. Ninth grade is also the grade of PISA test's participants, which makes the results can also work as an evaluation of Indonesia's low PISA results.

Instruments

To answer the first research question, this study used a mathematical assignment from the International Centre for Science, Technology, Engineering, and Math Education (ICSE, 2018) with a slight adjustment to suit the Indonesian context, which is summarized in [Figure 1](#). The assignment was an open-ended mathematical problem with real-life settings whose characteristics are similar to the HOT features put forward by Resnick (1987) such as non-algorithmic, allow multiple solutions, involve uncertainty, and encourage self-regulation of the thinking process. The use of open-ended problems supports the development of creative and critical thinking which is core to HOT (Emilya et al., 2010). A working sheet was provided to help students structuring their answers.

A typical situation at the checkouts at Indonesian supermarkets: You get 50-cents coin as change, and it then fills up your wallet. The founders of the donation campaign “Germany rounds up” had a brilliant idea: Customers can donate the amount of change they get from rounding it up by saying ‘round-up please’ at the checkout. If the campaign is applied in Indonesia where customers can donate the amount of change they get from rounding it up to the nearest hundred, **how much money can be collected in an average supermarket in a day?** First, think about how many customers shop there each day and how many would decide to donate their rounded change? Ensure that your arguments and calculations can be the basis of consideration **whether this campaign should be applied or not**. You can use the Internet or other media to find information.

Figure 1. The overview of the ICSE task

To categorize students' level of HOT skills, the rubric adapted from Exemplars (2014), WSU (2001), AACU (2009), and Alfrey and Cooney (2009) was utilized (see [Table 1](#)). To answer the second and third research questions, this study used data from semi-structured interviews with the teachers. The interview questions reflect the research focus of the aspect of curriculum, pedagogy, and assessment, as the essential factors in the development of HOT skills (Abdullah et al., 2017), and the protocol is adjustable according to the perception of the interviewer about what is necessary and relevant (Robson, 2002).

Table 1. Adapted HOT rubric

Category	Level	Description	Example
Explanation of issues	1	Issue/problem to be considered critically is not stated or interpreted incorrectly.	<ul style="list-style-type: none"> • The price in the supermarket should be rounded down.
	2	Issue/problem to be considered critically is stated but description leaves ambiguities, boundaries undetermined, and/or backgrounds unknown.	<ul style="list-style-type: none"> • The change from the supermarket can be rounded up.
	3	Issue/problem to be considered critically is stated, described, and clarified so that understanding is not seriously impeded by omissions.	<ul style="list-style-type: none"> • The change from the supermarket can be rounded up to the nearest hundred.
	4	Issue/problem to be considered critically is stated clearly and described comprehensively, delivering all relevant information necessary for full understanding.	<ul style="list-style-type: none"> • The change from the supermarket can be rounded up to the nearest hundred. The author wants to know how much money can be collected in average supermarket in a day and whether the campaign should be applied in Indonesia.
Reasoning	1	Arguments are made with no mathematical basis. No sound reasoning nor justification for reasoning is present.	<ul style="list-style-type: none"> • 267 million people in Indonesia would donate 50 cents a day.
	2	Arguments are made with some mathematical basis. Some correct reasoning or justification for reasoning is present.	<ul style="list-style-type: none"> • Out of 1000 customers come to the supermarket, 60% would be willing to donate.
	3	Arguments are constructed with adequate mathematical basis. A systematic approach and/or justification of correct reasoning is present.	<ul style="list-style-type: none"> • Out of 1000 customers come to the supermarket, 60% would be willing to donate because 20% is poor, 10% have rounded payment, 5% reject to donate, and 5% pay with debit card.
	4	Deductive arguments are used to justify decisions and may result in formal proofs.	<ul style="list-style-type: none"> • According to the graph in the study of donation campaign (i.e., https://www.slideshare.net/YayasanLembagaKonsum/praktek-penggalangan-dana-publik-kajian-terhadap-penerapan-prinsip-transparansi-dan-akuntabilitas), 72% of customers' decision to donate lay on the credibility of the company. Therefore, _____

Category	Level	Description	Example
Evidence	1	Does not identify or assess the quality of supporting evidence.	<ul style="list-style-type: none"> • There are 100 customers in a day.
	2	Merely repeats information provided. Does not justify position or distinguish between fact and opinion	<ul style="list-style-type: none"> • According to a study, in average there are 87 customers in a day.
	3	Examines evidence and questions the quality. Distinguishes between fact and opinion.	<ul style="list-style-type: none"> • My answer is not accurate because of the limited resources that I used.
	4	Provides a well-developed examination of the evidence and questions its accuracy, relevance, and completeness. Clearly distinguishes between fact and opinion.	<ul style="list-style-type: none"> • The data was found from (source citation). The data, nevertheless, was based on a survey in 2012 so it could not accurately reflect situation that happen at the moment. Therefore, ____
Problem-solving	1	No strategy is chosen, or a strategy is chosen that will not lead to a solution.	<ul style="list-style-type: none"> • The total collected donation is 129 billion.
	2	A partially correct strategy is chosen, or a correct strategy for only solving part of the task is chosen.	<ul style="list-style-type: none"> • The total collected donation = estimated number of customers x total payment.
	3	A correct strategy is chosen based on the mathematical situation in the task.	<ul style="list-style-type: none"> • The total collected donation = estimated number of customers x 50 cents.
	4	An efficient strategy is chosen and progress towards a solution is evaluated.	<ul style="list-style-type: none"> • Information about the results of donations of a supermarket brand for a certain duration divided by the duration of time to get many donations per day as an estimate.

Category	Level	Description	Example
Conclusions and evaluation	1	Attempts a conclusion or evaluation that is illogical or inconsistent with evidence presented, or omits a conclusion or solution altogether.	<ul style="list-style-type: none"> The campaign should not be applied because it can disadvantage the supermarket.
	2	Presents abbreviated or simple conclusions that are mostly consistent with evidence presented, with minor inconsistencies or omissions.	<ul style="list-style-type: none"> Donation campaign must be applied as an example for our generation to help each other.
	3	Clearly states and discusses conclusions. Organizes a conclusion or solution that is complete, logical, and consistent with evidence presented.	<ul style="list-style-type: none"> If the collected donation in a day is xx IDR, then we can collect xx IDR in a year. Since the amount is quite big, we should apply the campaign to help more people in needs.
	4	Clearly states and discusses conclusions. Considers implications and consequences of the conclusion in context, relative to assumptions, and supporting evidence. Provides reflective thought with regards to the assertions.	<ul style="list-style-type: none"> If the collected donation in a day is xx IDR, then we can collect xx IDR in a year. The campaign, therefore, should be implemented because the amount of money is quite big so it can help more people in needs. However, the campaign can be effective or ineffective. To make it effective, _____

Procedure and data collection

The three different schools from each city were selected based on stratified random sampling. First, 1,055 schools in Jakarta and 199 schools in Palembang were classified into three categories according to their average mathematics scores in the 2018 national examination, which are 'High' (average score > 70), 'Medium' ($40 < \text{average score} < 70$), and 'Low' (average score < 40). This is to avoid convoluting the results because a recent study has found that academic achievement has a positive correlation with the level of HOT skills (Tanujaya, Mumu, & Margono, 2017). By having such categorization, the study was expected to cover all three different layers of HOT skills' level among teachers and students. Second, the five schools in each category were selected randomly by using Microsoft Excel. The consideration of choosing five schools was made for a conservative purpose, with the consideration that it was not easy to get fast responses and permission from schools to carry out the study. Third, all thirty schools were contacted to see which schools responded positively. The target of one school per category is achieved. Fourth, a schedule was arranged with the ninth-grade mathematics teachers from each school and permission was asked to provide accessibility for the students to access the internet when working on the ICSE task either from their phones or computers.

The ICSE task for the students took approximately 60 minutes in total, beginning with a short introduction about the study. Students were then given 5 minutes to read the problem. After which, two examples of price catalogs from two supermarkets were shown to help students visualize the problem. They were allowed to ask questions, but only to check understanding. No further clues were given. Students were also permitted to use their cell phones or laptops with continuous supervision to ensure there was no misuse. In the end, they submitted both the ICSE task papers and the worksheets to prevent the distribution of the materials to other classes.

The collected data in this research included the voice recordings of teachers' semi-structured interviews and the students' worksheets of the ICSE task. The face-to-face interviews with teachers took place at their schools. It started with a brief introduction to the study and continued with the questions, which took approximately 30 minutes for each and were tape-recorded with consent.

Data analysis

The units of analysis in this study are a group of ninth-grade students and a group of ninth-grade mathematics teachers. To evaluate students' ability to demonstrate HOT skills, the results of the ICSE task were analysed by the researcher with a rubric that had been initially developed. The rubric values students' abilities to 1) explain the issues, 2) justify the reasoning, 3) evaluate evidence, 4) solve problems, and 5) state reflective conclusion and evaluation, in which each category ranges from Level 1 to Level 4. A statistical representation of students' level of HOT skills in mathematics was created with respect to their score levels in each category. Since the students answered open-ended questions, the categorization process was also done by a second coder. Interrater reliability was assessed by

calculating the percentage of agreement. Lombard et al. (2010) believe that around 10% of the total data should be adequate for the inter-rater test. We took 10% of the total students' worksheets from each school by numbering the worksheets and selecting them using Google Random Number Generator. The interrater test led to an agreement of above 90% for all categories, which is considered an acceptable inter-rater agreement (Stemler, 2004).

To investigate teachers' views of HOT skills, the interviews were transcribed. Thematic analysis was conducted using an inductive approach. The inductive approach assigned themes or categories. The answers to the interview questions related to curriculum, pedagogy, and assessment were summarized and classified into three to four subcategories by identifying and coding quotations that reflected different views.

To understand teachers' challenges, needs, and existing supportive factors in developing HOT skills, the interviews were transcribed. Thematic analysis was conducted using a deductive approach, which involved coding all transcriptions and assigning themes that were observed in previous research (Alhassora, Abu, & Abdullah, 2017). All the excerpts mentioned in the result section were translated into English.

RESULTS AND DISCUSSION

This section presents the results and analysis of students' HOT assessment which consisted of 372 students' worksheets, as well as the thematic analysis of one-on-one interviews with six selected ninth-grade mathematics teachers from six secondary schools in Jakarta and Palembang.

Students' ability to demonstrate HOT skills

The results of students' HOT assessment are presented in Table 2. Overall, around half of the students were able to achieve Level-3 in Problem-solving (43%) and Conclusion and evaluation (53%). However, students most likely experienced difficulty in Reasoning and Evidence, with only 1% and 3% in Level-3 respectively, and 0% in Level-4. More detailed findings in each category are as reported in Table 2.

Table 2. Results of students' HOT assessment

Category	Number of students				
	No answer	Level 1	Level 2	Level 3	Level 4
Explanation of issues	87 (23%)	87 (23%)	73 (20%)	101 (27%)	24 (6%)
Reasoning	20 (5%)	298 (80%)	52 (14%)	2 (1%)	0 (0%)
Evidence	20 (5%)	322 (87%)	19 (5%)	11 (3%)	0 (0%)
Problem-solving	20 (5%)	133 (36%)	58 (16%)	159 (43%)	2 (1%)
Conclusion and evaluation	17 (5%)	144 (39%)	11 (3%)	198 (53%)	2 (1%)

Explanation of issues

This category values students' capability to identify and describe the main problem, to deliver all relevant information for full understanding. It also connects to 'comprehension' of the task and includes some elements of horizontal mathematization where students need to understand what the 'mathematical' question/task is (Anwar et al., 2012) and to use mathematical tools to organize and solve problems situated in real-life situations (Van den Heuvel-Panhuizen & Drijvers, 2013). In the ICSE task, the problem was presented in a written text, so it was necessary to decode the text information to understand the mathematical issues in the problem situation (Sbaragli & Franchini, 2017). Students were expected to point out the relevant or notable information to have a correct understanding of the concept of 'rounding-up to the nearest hundred,' and to formulate the mathematical question being asked. The given information was not enough to start a calculation, so students needed to make estimations before they could work out the solutions.

The results in [Table 2](#) show that 23% of the students did not identify the issue despite being instructed by the question. They jumped directly to a mathematical formula to get the answer to the question about the total collected donation per day. The same number of students, 23%, were at Level-1. These students tried to formulate issues, but misinterpreted the problems and/or were unable to describe the situation correctly. For instance, one of the students specified the issue as "how to donate at least 50 cents" instead of "finding out the average collected donation in a day" (see [Figure 2](#)).

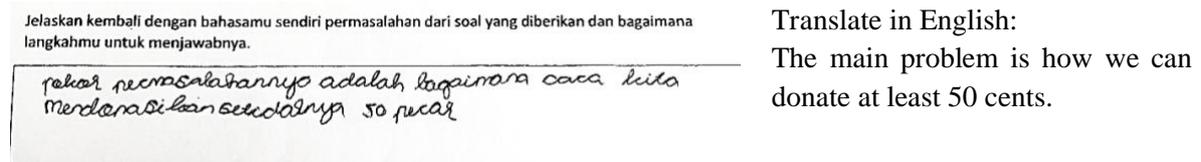


Figure 2. An example of students' answers at Level-1 in Explanation of issues

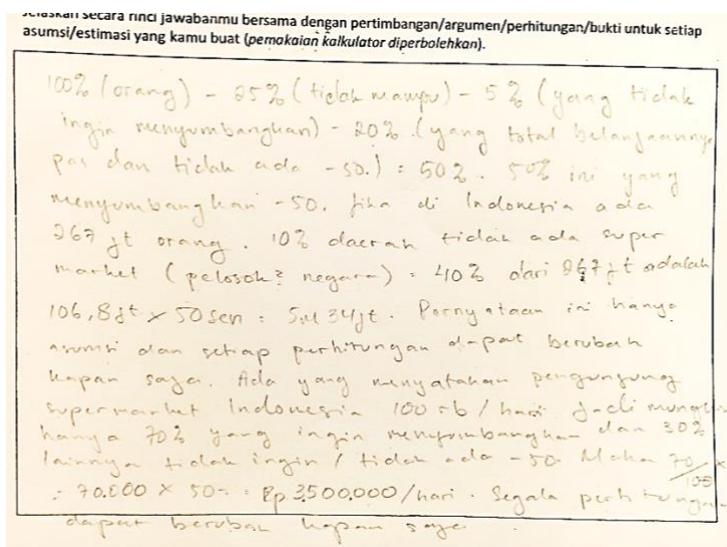
This misunderstanding could lead to irrelevant answers such as "having a charity box." The issue to be considered critically was not reported. Furthermore, around 20% of the students were able to highlight the concept of rounding up but leaving ambiguities. It was not clear whether the rounding was done to the nearest hundred or nearest thousand. About 27% of the students managed to summarize the main point of the passage correctly and applied the mathematical concept of 'rounding up to the nearest hundred.' The remaining 6% of the students successfully delivered all relevant information and formulated the mathematical question.

Reasoning

The reasoning category evaluates students' capability to build mathematical argumentations to justify their claims. In the ICSE task, students are expected to be able to utilize proportional reasoning – use ratios, proportions, rates, and percentages to explain, draw conclusions, or make decisions on the

estimated number of customers who would donate. They were expected to provide “why and how” such estimates were chosen. The mathematical argumentations should be clear, logical, and well explained.

The results in [Table 2](#) show that 5% of the students left the sheet blank. Most of the students, about 80%, stood at Level-1. These students did not make any proportional reasoning or other calculations to estimate the number of customers who would make donations. There was also no statement whether this was done because they assumed that all customers would not decline. The number of people who donated, therefore, was always equal to the number of customers. About 14% of the students did proportional reasoning by considering, for instance, that only 60% of customers would be willing to donate. Only 1% of the students did proportional reasoning with several considerations that were constructed with an adequate mathematical basis. For example, one of the students justified the chosen estimate by considering several aspects such as 1) the finance of the customers (e.g., rich or poor), 2) the willingness of the customers, 3) not all payments need a rounding, and 4) not everyone had access to the supermarket (see [Figure 3](#)). One student considers the possibility of paying by credit or debit card instead of cash. If customers pay by card, then the cashier will not ask for a rounding.



Translate in English:
 100% (people) – 25% (underprivileged) – 5% (do not want to donate) – 20% (have rounded total amount) = 50%. This 50% are those who donate 50 cents. In Indonesia, there is 267 million people. 10% of the regions are villages with no supermarket (remote areas) = 40% of 267 million is 106.8 million x 50 cents = 5.34 billion. This is just an assumption and every calculation can change anytime.

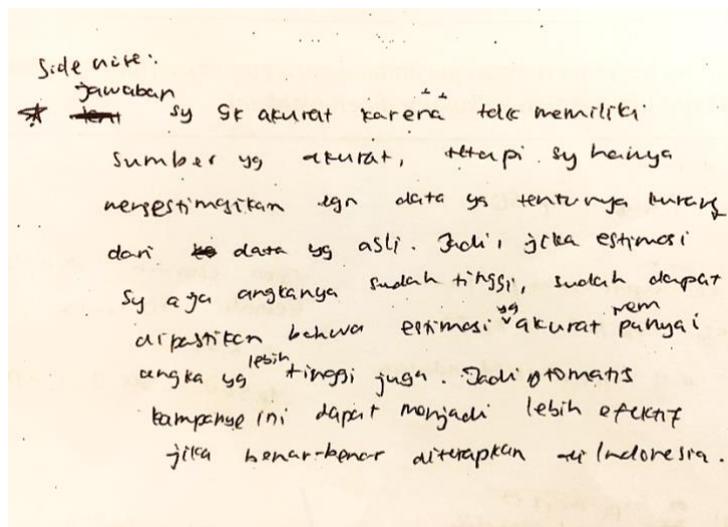
Figure 3. An example of students' answers at Level-3 in Reasoning

Evidence

The evidence category measures students’ capability to critically examine information or data that they found and questions its accuracy, relevance, and completeness. Because students were given opportunities to search for information on the Internet or other resources, they were expected to be able to assess whether the data source was valid or not, to distinguish between fact and opinion and not simply trust information based on ‘appeal to authority’ (because a prominent figure said so, then it’s true) or ‘social acceptability’ (because everyone believes it, then it’s true). Overgeneralizing is also one of the common errors that make generalization of the whole group from one or very few examples

(Brookhart, 2010). In the ICSE task, the evidence refers to the data that supports the claim about the number of customers who came to the supermarket each day.

The results in [Table 2](#) show that 5% of the students left the sheet blank. In line with Reasoning, the highest weight in Evidence stood at Level 1. About 87% of the students made direct claims that were simply based on their opinions. For example, they assumed that ‘there are 100 customers who come to a supermarket in a day’ without any proof or argumentation. Around 5% of the students at Level-2 mentioned that they found the information from a study or a website. The credibility of such information could not be assessed because the information was incomplete. It was unclear whether the data was a fact or an opinion. The remaining 3% based their claims on data taken from a valid source with proper citation; or examined the accuracy of their claims. For instance, one of the students explicitly wrote a remark on the accuracy of the estimation (see [Figure 4](#)). However, none of the students was able to perform elaborate examination of the evidence and question its accuracy, relevance, and completeness.



Translate in English:

My answer is not accurate because (I) don't have an accurate source, but I only estimate with the data that is obviously less than the actual data. So, if my estimation has already had a high number, it is certain that an accurate estimate would have a high number too. So, automatically, the campaign would be more effective if it is actually implemented in Indonesia.

Figure 4. An example of students' answers at Level-3 in Evidence

Problem-solving

The problem-solving category assesses students' abilities to construct a mathematical model that describes the essence of the elements and relations involved in a particular situation that is of interest. They had to choose a correct and efficient strategy based on the given mathematical situation. In the ICSE task, students were expected to formulate a mathematical model to find the total amount of collected donation in a day.

The results in [Table 2](#) shows that 5% of the students left the sheet blank. About 36% of the students at Level-1 either stated the answer without any calculations, focused on the social aspects instead of answering mathematically, or made an irrelevant model. For example, one of the students assumed the amount of donation as 1% of people's daily salaries (see [Figure 5](#)). This did not match with

the given situation to calculate average donation in a supermarket based on the rounding concept. Moreover, some students calculated the total income received by the supermarket instead of the total donation. They multiplied the number of customers by the total payment made. Around 16% of the students were able to choose a correct strategy but calculate incorrectly and/or understood the rounding-up concept partially. Some of them did the round-up to the nearest thousand instead of the nearest hundred. Nearly half of the students, 43%, were able to build a correct mathematical model and performed the right calculation, but only 1% came up with an efficient strategy to find the estimated amount of donation in a day. These 1% students searched for information on donation programs that had already been run by a supermarket brand, examined its total collected donation, and calculated the mean to get the amount of donation per day. This was not a strategy being suggested by the questions in the task.

~~Jumlah customer tiap hari bisa bervariasi 100-500 orang.~~
 Jumlah customer hypermart mall WTC
~~sekitar~~ rata-rata perhari adalah ~~100~~ 170 orang.
 Rata-rata pengeluaran orang Jakarta sekitar
 2 juta perbulan.
 total yang bisa didonasikan (sebelum round up)
 mungkin sekitar dika
 sebulan 2 juta, 1 hari sekitar Rp. 60.000
~~jumlah~~ jumlah yg bisa didonasikan kira-kira
 1% perhari → 600 perak
 perbulan: Rp. 18.000
 pertahun: Rp. 216.000
 Perhari 600 x 170 orang = Rp. 102.000

Translate in English:
 The average number of customers in Hypermart WTC Mall in a day is 170 people. The average spending of a Jakarta resident is about 2 million per month. The total (money) that can be donated (before rounding up) is approximately:
 if in a month (they receive) 2 million, 1 day is about 60,000 IDR. The amount that can be donated approximately 1% per day → 600 cents.
 per month: 18,000 IDR
 per year: 216,000 IDR
 per day: 600 IDR x 170 people = 102,000 IDR

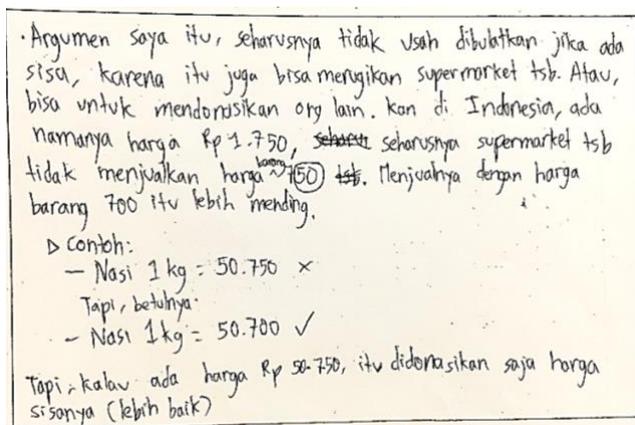
Figure 5. An example of students’ answers at Level-1 in Problem-solving

Conclusion and evaluation

The conclusion and evaluation category evaluates students’ capability to critically discuss the conclusion and implications and consequences of the solving in a context, as well as to provide reflections of the assertions. In the ICSE task, students were expected to state the conclusion based on the evidence or the presented calculation and to consider the implications and consequences of the conclusion in a context.

The results in [Table 2](#) show that about 5% of the students left the sheet empty. About 39% of the students did not state any conclusion or attempted to draw conclusions but illogical or inconsistent with the evidence being presented. For instance, one student argued that the price should not be rounded up because it could disadvantage the supermarket (see [Figure 6](#)). The campaign costs did not come from the supermarket so the argument was not logical. Around 3% students provided a conclusion with minor inconsistencies. These refer to conclusions that were in line with the calculation results, but they made

the performed calculation meaningless. For example, these students concluded that the donation campaign should be implemented because “people have to share” or “people have to do good,” so the conclusions were not associated with the mathematical calculation. More than half of the students, 53%, reached level-3 by clearly stating the conclusions that are connected with the calculation presented. About 1% provided reflective thought by considering the implications and consequences of the conclusion they made. One student suggested creating a system in the computer to prevent corruption because there was a possibility that the money would be taken by the cashier or the officer.



Translate in English:

My argument, it should not be rounded if there is a change because it can also disadvantage the supermarket. Or, (it) can be used for donation. In Indonesia, there are (things with) price 1,750 IDR. Ideally, the supermarket should not sell things with (ending) price 750. It is better to sell it with (ending) price 700. For example:
 Rice 1 kg = 50,750 ×
 Rice 1 kg = 50,700 ✓
 But, if there is a thing with price 50,750 IDR, just donate the change (better).

Figure 6. An example of students’ answers at Level-1 in Conclusion and evaluation

The results of students’ HOT assessment indicate that one of the major problems is in the transfer of knowledge. This is indicated by a high percentage of students (>80%) being in Level-1 Reasoning. Students with the ability to transfer their knowledge would be able to use contextualized reasoning to solve various problems in outside of school settings, such as the supermarket (Barrouillet & Gauffroy, 2013). Almost all students assumed all customers would be willing to donate, all customers would have the unrounded price to pay, and all customers would pay by cash. There was no justification found for the assumptions made, although justification is a key aspect of adaptive reasoning (Kilpatrick et al., 2001).

The other major problem lies in critical thinking, which also relates to information literacy skills. This is indicated by the high number of students falling under Level-1 Evidence. Students tended to use information without examining its quality; without distinguishing between fact and opinion and its relevance to the problem in the context. This finding is also aligned with Wijaya (2016) who found that Indonesian students did not possess three characteristics of information literacy; i.e., recognizing information needs, locating and evaluating the quality of information, and making effective and ethical use of information. The lack of information literacy skills might be because of less engagement with technology, but Gibson (2012) argued that “exposure to technology does not automatically equate to proficiency in technology.” Students might also find it challenging to locate information to meet their

needs, come up with an effective word search, infer useful links within search results, and scan for relevant information within websites (Leu et al., 2011).

These findings, nevertheless, are slightly different from Hadi et al. (2018), who found that crucial difficulties experienced by the students in solving HOT test problems were process skills and transformation errors. Process skills errors are marked by students’ errors in math calculation, whereas transformation errors are those when developing a mathematical model. In this case, only 16% showed process skills errors and nearly 60% of the students were able to transform the problem into a correct mathematical model. However, some of the models were not perfectly constructed due to the lack of transfer knowledge skills and critical thinking skills. Several aspects or variables were not considered in constructing the mathematical model.

Teachers’ views of HOT skills in terms of curriculum, pedagogy, and assessment

Curriculum

The interview questions related to the curriculum were divided into two categories, which explored the “concept of HOT skills” and the “purpose of developing HOT skills.” The results showed that the teachers’ answers with respect to the concept of HOT skills could be categorized into HOT skills as: a) critical and creative thinking skills; b) flexible problem-solving skills; and c) skills of talented or higher-ability students. Meanwhile, the teachers developed HOT skills with the purpose of: a) selecting students based on their mathematical abilities; b) preparing students to tackle HOT problems in the national examination; and c) preparing students for society. The illustrative quotations of each subcategory can be seen in Table 3.

Table 3. Teachers’ views of HOT skills in terms of curriculum

Category	Subcategory	Sample quotations	Interviewee
Concept of HOT skills	A) Critical and creative thinking skills	“Students with HOT will be able to think critically, creatively, be able to analyze, compare, create, not just to memorize.”	Teacher 3
			Teacher 2
			Teacher 6
	B) Flexible problem-solving skills	“One’s ability to solve mathematical problems with a variety of strategies and methods; methods that are not rigid or conventional.”	Teacher 5
			Teacher 6
	C) Skills associated with talented or higher-ability students	“50:50 in determining whether HOT is a skill that could be developed only by gifted students. Sometimes, (HOT) does not give opportunities for students with lower abilities. I hope that HOT can be developed by all students, but it is easier to be	Teacher 1
			Teacher 2
			Teacher 4

Category	Subcategory	Sample quotations	Interviewee
		developed by students who have higher abilities than others.”	
Purpose of developing HOT skills	A) Selecting students	“Assessing HOT skills is important to sort out which students have higher-order and which students have lower-order.”	Teacher 2
	B) Preparing for the national examination	“Yes (it is important), so that if there are HOT problems in the national examination, students would be able to solve them. By giving a HOT problem like what had just been done, students will know what HOT problem looks like.”	Teacher 1
	C) Preparing for society	“There are definitely benefits. Mathematics is the mother of all sciences. It is a thinking tool instead of a discipline. So, if the level of thinking has reached the HOT level, it will really help her/him tackling other subjects or problems that require analytical skills and problem-solving strategies.”	Teacher 3 Teacher 4 Teacher 5 Teacher 6

The results indicate that there was a belief that HOT skills in mathematics can only be developed by talented students or students with higher mathematical abilities. The teachers believed that the cognitive demands of HOT tasks were beyond the capabilities of low-achieving students. This belief cast doubt among teachers that low-achieving students would be able to develop HOT skills. As a result, the low-achieving students may experience lower-order instructional emphasis because teachers see these students as ‘stuck’ at an early phase of the learning process (Raudenbush, Rowan, & Cheong, 1993).

While the previous study concluded that teachers in Indonesia were already aware of the importance of HOT skills (Retnawati et al., 2018), the results demonstrated that there were teachers who developed HOT skills in the classroom to distinguish between students with higher-order thinking and students with lower-order and to simply train students so that they pass the national examination. According to Kirkpatrick and Zang (2011), exam-oriented education can restrain a student’s imagination, creativity, and sense of self, which are vital qualities for a child’s ultimate success at school and society.

Pedagogy

The interview questions related to the pedagogy focus on exploring teachers’ knowledge of the “teaching strategies to develop HOT skills.” The answers could be grouped into three subcategories, namely: a) teachers who consider student-centered pedagogy (e.g., discussion group, problem-based

learning, and hands-on learning); b) context-based learning; and c) the appropriateness or suitability of the learning environment. The illustrative quotations of each subcategory can be seen in [Table 4](#).

Table 4. Teachers’ views of HOT skills in terms of pedagogy

Category	Subcategory	Sample quotations	Interviewee
Teaching strategies to develop HOT skills	A) Student-centred pedagogy	“Discussions and Q&A help students develop HOT. One way to develop it is by giving stimulus such as: ‘Why is it like this?’”	Teacher 1 Teacher 6 Teacher 4
	B) Context-based learning	“PMRI (<i>Indonesian Realistic Mathematics Education</i>) is a good approach.”	Teacher 5
	C) Suitable learning environment	“Firstly, consider the learning environment. We cannot teach HOT material in the afternoon session. Secondly, utilize smart students. We brief them so they can teach their peers. However, it usually is a waste of time if the number of smart students in the classroom is small.”	Teacher 2 Teacher 3

Assessment

The interview questions related to the assessment evaluates teachers’ perception of the “characteristics of a HOT problem” and “how they evaluate ICSE task as a HOT problem.” The answers could be placed into four subcategories, which are teachers who recognize or identify HOT problems as: a) contextual problems; b) difficult problems; c) problems that requires multiple steps; and d) problems that are long, unfamiliar, and have pictures in it. The illustrative quotations of each subcategory can be seen in [Table 5](#).

Table 5. Teachers’ views of HOT skills in terms of assessment

Category	Subcategory	Quotations	Interviewee
Characteristics of a HOT problem	A) Contextual	“HOT problem must be contextual. If the problem is contextual, then the problem solving will be more logical or more reasonable, because mathematics is a subject that is related to rational matters.”	Teacher 1 Teacher 5 Teacher 6
	B) Difficult	“HOT problem is identical with a very difficult problem. From the term itself, it stands for ‘higher-order thinking.’”	Teacher 1

Category	Subcategory	Quotations	Interviewee
	C) Require multiple steps	“HOT problem does not have to be context-based. For example, direct problem (not in the form of a storyline) with a higher level, not level 1 but level 3, for instance. Level 3 means that the problem cannot be solved by one step only.”	Teacher 2 Teacher 3 Teacher 4
	D) Long, unfamiliar, include pictures	“(to identify the HOT problem in the national examination), the HOT problem is usually long, has never been given before, and has pictures in it.”	Teacher 3

The results show that half of the teachers identified HOT problems as contextual problems, but there were also other views of it. One of the teachers believed that HOT problems were identical with difficult problems. According to Sydoruk (2018), “Difficulty refers mainly to the amount of effort, cognitive or physical, that a student needs to exert to complete a task but does not account for the ways that a student must think about the task or problem in order to solve it.” The level of difficulty, therefore, is adjustable based on students’ learning style, pre-knowledge, and personal comfort level with the problem (Bieri & Blacker, 1956). There was also a view that the existence of ‘long passage’ and ‘picture’ are indicators of a HOT problem. This belief was used to distinguish between HOT and LOT problems in the national examinations. This also reflects one of the listed misconceptions by Nugroho (2018) that a phenomenon, a case, or an event that reflects HOT problem should be presented in a long storyline.

Teachers’ challenges, needs, and existing supportive factors in developing HOT skills

In the interview, there were teachers who felt that they had never applied HOT, occasionally applied HOT, and always applied HOT. To understand the factors that affect how HOT is implemented, this study classifies these teachers’ challenges, needs, and the existing supportive factors based on their frequency in addressing HOT skills (see Table 6).

Table 6. Teachers’ challenges, needs, and existing supportive factors in developing HOT skills

Frequency of HOT use	Challenges	Needs	Existing supportive factors
Never	<ul style="list-style-type: none"> • Need to focus on passing the national exam • Classroom diversity • Lack of knowledge of the HOT problem characteristics 	<ul style="list-style-type: none"> • Guidebook and support from the government to understand HOT • Equality between private schools and public schools 	

Frequency of HOT use	Challenges	Needs	Existing supportive factors
Occasionally	<ul style="list-style-type: none"> • Lack of mathematical literacy and reading skills among students • Lack of motivation among students • Time-consuming (complex preparation and implementation) • Too many administrative requirements and competency demands in the curriculum • Need to focus on passing the national exam 	<p>in receiving information</p> <ul style="list-style-type: none"> • Teacher training with practical activities • Guidebook • Colleague-support • Long-term teacher training 	<ul style="list-style-type: none"> • Knowledge of Indonesian Realistic Mathematics Education (PMRI) • Familiarization of PISA test
Always	<ul style="list-style-type: none"> • Classroom diversity • Lack of motivation among students 	<ul style="list-style-type: none"> • Teacher training 	<ul style="list-style-type: none"> • Colleague-support

The results suggest that the knowledge of existing teaching strategies to develop HOT, HOT sample problems, and colleague-support influence the development of HOT skills in the classroom. Teacher training should be available for all teachers, both in public and private schools, without any prioritization. Wolgast and Fischer (2017), moreover, stated that teaching profession may induce stress due to the time constraints, heavy workload, and extra-curricular obligations. They believed that colleague-support served as a resource for teachers and that it had a positive effect on their performance.

Furthermore, the administrative duties in the national curriculum tend to hinder the development of HOT skills. Teachers had limited time to learn about HOT skills as a ‘new material’ due to the heavy workloads. Werang (2018) argued that the amount of teachers’ workload has a significant positive correlation to their burn out, which could bring a negative impact on students’ academic performance (Gwambombo, 2013).

CONCLUSIONS

This study aimed to examine HOT in Indonesian lower secondary mathematics classrooms by exploring teachers’ views and assessing students’ ability to demonstrate HOT skills in mathematics.

The findings show that most of the students could construct a mathematical model but experiencing difficulty in applying knowledge into new contexts and in applying creative thinking. The other major problem lies in critical thinking, which is indicated by the lack of information literacy skills.

In terms of curriculum, some of the teachers were familiar with the concept of HOT skills, but there were views that teaching HOT skills was suitable only for talented or students with higher intelligence. There was also a practice of exam-oriented education which can discourage students' creativity. In terms of pedagogy, half of the teachers were able to mention appropriate teaching strategies to develop HOT, but some of the answers were normative. Inconsistency was also found in the teachers' responses when explaining the strategy. In terms of assessment, half of the teachers identified HOT problems as contextual problems. Some teachers associated HOT problems with difficulty and with a long story-line.

This study, furthermore, found that the factors that influence the development of HOT skills are the knowledge of existing teaching strategies to develop HOT, the familiarity with HOT problems, and colleague-support. Besides that, the amounts of administrative duties in the national curriculum may also hinder the development of HOT skills. These findings raise several implications. We need a more detailed outline of levels of higher order thinking in mathematics that is consistent and supported with resources and assessment methods. Teachers need support in terms of teacher training and guidebooks and this should be equally distributed to both public and private schools to ensure a common view of HOT skills, its learning strategy, and characteristics of the assessment. Finally, schools can support colleagues to work together in (inter)disciplinary teams on the conceptualization and implementation of HOT in their teaching practices.

The findings of our study, nonetheless, should be taken with precaution because there are some limitations that need to be taken into account. Even though we ensured the diversity of schools in our sample, the generalizability of the study is limited due to the small sample size compared to the number of lower secondary students and teachers in Indonesia. The use of the ICSE task did not allow us to evaluate the full richness of all HOT dimensions, and teacher reports' provide a limited view on real classroom practice. The benchmark for comparison is also limited due to the limited prior research in Indonesian contexts.

Developing higher order thinking skills in mathematics is important in our current quickly changing and technological society. Policy makers are aware of this importance. This study shows that still some steps are needed for a sustainable implementation of higher order thinking in lower secondary mathematics classrooms.

ACKNOWLEDGMENTS

The authors would like to thank *Lembaga Pengelola Dana Pendidikan (LPDP)* for their financial support, as well as participated students and teachers for their participation in this study.

REFERENCES

- Abdullah, A. H., Mokhtar, M., Halim, N. D. A., Ali, D. F., Tahir, L. M., & Kohar, U. H. A. (2017). Mathematics teachers' level of knowledge and practice on the implementation of higher-order thinking skills (HOTS). *EURASIA Journal of Mathematics Science and Technology Education*, 13(1), 3-17. <https://doi.org/10.12973/eurasia.2017.00601a>.
- Ahmad, S., Prahmana, R. C. I., Kenedi, A. K., Helsa, Y., Arianil, Y., & Zainil, M. (2018). The instruments of higher order thinking skills. *Journal of Physics: Conference Series*, 943(1), 012053. <https://doi.org/10.1088/1742-6596/943/1/012053>.
- Alhassora, N. S. A., Abu, M. S., & Abdullah, A. H. (2017). Inculcating higher-order thinking skills in mathematics: Why is it so hard? *Man in India*, 97(13), 51-62. <https://doi.org/10.2478/v10274-012-0006-7>.
- Aini, N. (2018, April 25). KPAI criticizes the use of HOTS problems in national exams [in Bahasa]. *Republika*. Retrieved from <https://republika.co.id/berita/pendidikan/education/18/04/25/p7qrd0382-kpai-kritik-penggunaan-soal-hots-dalam-ujian-nasional>.
- Alfrey, K., & Cooney, E. (2009, June). *Developing a rubric to assess critical thinking in assignments with an open-ended component*. Paper presented at the American Society for Engineering Education Annual Conference and Exposition, Austin, TX.
- Anwar, L., Budayasa, I. K., Amin, S., & De Haan, D. (2012). Eliciting mathematical thinking of students through realistic mathematics education. *Journal on Mathematics Education*, 3(1), 55-70. <https://doi.org/10.22342/jme.3.1.620.55-70>.
- Apino, E., & Retnawati, H. (2017). Developing instructional design to improve mathematical higher-order thinking skills of students. *Journal of Physics: Conference Series*, 812(1), 012100. <https://doi.org/10.1088/1742-6596/812/1/012100>.
- Ariyana, Y., Pudjiastuti, A., Bestary, R., & Zamroni. (2018). *The Learning Handbook Oriented Higher-Order Thinking Skills* [in Bahasa]. Jakarta: Direktorat Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan.
- Association of American Colleges and Universities (AACU). (2009). *Critical thinking VALUE rubric*. Retrieved from <https://www.aacu.org/value/rubrics/critical-thinking>.
- Barrouillet, P., & Gauffroy, C. (2013). *The Development of Thinking and Reasoning*. London: Psychology Press.
- Bieri, J., & Blacker, E. (1956). The generality of cognitive complexity in the perception of people and inkblots. *Journal of Abnormal and Social Psychology*, 53, 112-117. <https://doi.org/10.1037/h0049167>.
- Bloom, B. (1956). *Taxonomy of Educational Objectives Handbook-I: The Cognitive Domain*. New York: David McKay.
- Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. Alexandria, VA: ASCD.
- Emilya, D., Darmawijoyo, & Putri, R. I. I. (2010). Developing open-ended problems for circle material to improve mathematical reasoning for eight grade students of Sekolah Menengah Pertama Negeri 10 Palembang [in Bahasa]. *Jurnal Pendidikan Matematika*, 4(2), 8-18. <https://doi.org/10.22342/jpm.4.2.316>.

- Exemplars. (2014). *Standards-Based Math Rubric*. Retrieved from https://exemplars.com/sites/default/files/2019-06/2019_exemplars_math_rubric_k5pdf.pdf.
- Gibson, T. (2012). On the interface: Between media cultures and worlds. *Religion & Literature*, 44(2), 179-188. Retrieved April 19, 2020, from www.jstor.org/stable/24397677.
- Gwambombo, I. (2013). *The Effect of Teachers' Workload on Students' Academic Performance in Community Secondary Schools: A Study of Mbeya City*. The Open University of Tanzania.
- Hadi, S., Retnawati, H., Munadi, S., Apino, E., & Wulandari, N. F. (2018). The difficulties of high school students in solving higher-order thinking skills problems. *Problems of Education in the 21st Century*, 76(4), 520-532. Retrieved from <http://oaji.net/articles/2017/457-1533495738.pdf>.
- Hantoro, J. (2018, April 14). The HOTS Problem that made the High School UN participants noisy [in Bahasa]. *Tempo*. Retrieved from <https://nasional.tempo.co/read/1079399/soal-hots-yang-bikin-gaduh-peserta-un-sma>.
- Husain, H., Bais, B., Hussain, A., & Samad, S. A. (2012). How to construct open ended questions. *Procedia Social and Behavioral Sciences*, 60, 456-462. <https://doi.org/10.1016/j.sbspro.2012.09.406>.
- International Centre for STEM Education (ICSE). (2018). *Quarterly problem: Round up, please!* Freiburg: University of Education. Retrieved from <https://icse.eu/quarterly-problem-mathematics/>.
- Kitcher, P. (1984). *The Nature of Mathematical Knowledge*. Oxford: Oxford University Press.
- Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding It Up: Helping Children Learn Mathematics*. Washington, DC: National Academy Press.
- Kirkpatrick, R., & Zang, Y. (2011). The negative influences of exam-oriented education on Chinese high school students: Backwash from classroom to child. *Language Testing in Asia*, 1(3), 36-45. <https://doi.org/10.1186/2229-0443-1-3-36>.
- Lakatos, I. (1976). *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge: Cambridge University Press.
- Leu, D. J., McVerry, J. G., O'Byrne, W. I., Kiili, C., Zawilinski, L., Everett-Cacopardo, H., Kennedy, C., & Forzani, E. (2011). The new literacies of online reading comprehension: Expanding the literacy and learning curriculum. *Journal of Adolescent & Adult Literacy*, 55(1), 5-14. <https://doi.org/10.1598/JAAL.55.1.1>.
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2010). *Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects*. Retrieved from http://matthewlombard.com/reliability/index_print.html.
- Mathematical Association of America (MAA). (1991). *A Call for Change: Recommendations for the Mathematical Preparation of Teachers of Mathematics*. Washington, DC: Author.
- National Council of Teachers of Mathematics. (1991). *Professional Standards for Teaching School Mathematics* (Working draft). Reston, VA: Author.
- National Research Council. (2001). *Classroom Assessment and the National Science Education Standards*. Washington, DC: National Academies Press.
- Nugroho, R. A. (2018). *Higher-Order Thinking Skills (HOTS)*. Jakarta: Grasindo.

- OECD. (2016). "PISA 2015 Results in Focus". *PISA in Focus*, No. 67. Paris: OECD Publishing. <https://doi.org/10.1787/aa9237e6-en>.
- OECD. (2018). "PISA for Development Mathematics Framework". *PISA for Development Assessment and Analytical Framework: Reading, Mathematics, and Science*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264305274-5-en>.
- OECD. (2019a). *Indonesia - Country Note - PISA 2018 Results*. Retrieved from https://www-oecd-org.proxy.library.uu.nl/pisa/publications/PISA2018_CN_IDN.pdf.
- OECD. (2019b). "What is PISA?". *PISA 2018 Assessment and Analytical Framework*. Paris: OECD Publishing. <https://doi.org/10.1787/2c7c311d-en>.
- Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal*, 30, 523-553. <https://doi.org/10.3102/00028312030003523>.
- Resnick, L. B. (1987). *Education and Learning to Think*. Washington DC: National Academy Press.
- Retnawati, H., Djudi, H., Kartianom, Apino, E., & Anazifa, R. D. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems of Education in the 21st century*, 76(2), 215-230. Retrieved from <http://oaji.net/articles/2017/457-1524597598.pdf>.
- Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers*. Oxford: Blackwell.
- Sbaragli, S., & Franchini, E. (2017). *Didactic evaluation of standardized fifth grade math tests* [in Italian]. Locarno: SUPSI, Dipartimento formazione e apprendimento.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for Research on Mathematics Teaching and Learning*, 334-370. New York: MacMillan.
- Smith, V. D., & Darvas, J. W. (2017). Encouraging student autonomy through higher order thinking skills. *Journal of Instructional Research*, 6, 29-34.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating inter-rater reliability. *Practical Assessment, Research, and Evaluation*, 9(4), 1-11. <https://doi.org/10.7275/96jp-xz07>.
- Suryadi, B. (2018). *The Implementation of HOTS problems model in national examinations need to be balanced with improvement of teacher and student ability* [in Bahasa]. Badan Standar Nasional Pendidikan. Retrieved from <http://bsnp-indonesia.org/2018/04/21/penerapan-soal-model-hots-dalam-ujian-nasional-perlu-diimbangi-dengan-peningkatan-kemampuan-guru-dan-siswa/>.
- Sydoruk, P. D. (2018). *An Analysis of the Higher Order Thinking Requirements of a Grade 8 Online-Based English Language Arts Skills Program*. Seton Hall University Dissertations and Theses (ETDs). Retrieved from <https://scholarship.shu.edu/dissertations/2495>.
- Tanujaya, B., Mumu, J., & Margono, G. (2017). The relationship between higher-order thinking skills and academic performance of student in mathematics instruction. *International Education Studies*, 10(11), 78-85. <https://doi.org/10.5539/ies.v10n11p78>.
- Van den Heuvel-Panhuizen, M., & Drijvers, P. (2013). Realistic Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education*. Dordrecht, Heidelberg, New York, London: Springer.

- Washington State University (WSU). (2001). *WSU Critical thinking Rubric*. Retrieved from https://assessment.trinity.duke.edu/sites/assessment.trinity.duke.edu/files/page-attachments/WashingtonStateUniversityCriticalThinkingProjectResourceGuide_000.pdf.
- Werang, B. R. (2018). The effect of workload, individual characteristics, and school climate on teachers' emotional exhaustion in elementary schools of Papua. *Cakrawala Pendidikan, 37*(3), 457-469. <https://doi.org/10.21831/cp.v38i3.20635>.
- Wijaya, A. (2016). Students' information literacy: A perspective from mathematical literacy. *Journal on Mathematics Education, 7*(2), 73-82. <https://doi.org/10.22342/jme.7.2.3532.73-82>.
- Wijaya, A. (2018). How do open-ended problems promote mathematical creativity? A reflection of bare mathematics problem and contextual problem. *Journal of Physics: Conference Series, 983*(1), 012114. <https://doi.org/10.1088/1742-6596/983/1/012114>.
- Wolgast, A., & Fischer, N. (2017). You are not alone: colleague support and goal-oriented cooperation as resources to reduce teachers' stress. *Social Psychology of Education, 20*(1), 97-114. <https://doi.org/10.1007/s11218-017-9366-1>.