International Journal of Evaluation and Research in Education (IJERE)
Vol. 12, No. 1, March 2023, pp. 468~476
ISSN: 2252-8822, DOI: 10.11591/ijere.v12i1.23831

Assessment of domain-specific prior knowledge: A development and validation of mathematical problem-solving test

Ijtihadi Kamilia Amalina[1], Tibor Vidákovich[2]
[1]Doctoral School of Education, University of Szeged, Szeged, Hungary
[2]Institute of Education, University of Szeged, Szeged, Hungary

Article Info

Article history:

Received Jan 31, 2022
Revised Nov 3, 2022
Accepted Nov 24, 2022

Keywords:

Domain-specific prior knowledge
Mathematical problem-solving
Psychometric analysis
Test

ABSTRACT

Science, technology, engineering, and mathematics (STEM) is a complex problem-solving that depends on the deep structures of domain-specific prior knowledge (DSPK) in mathematics. However, there is a lack of mathematics DSPK tests measuring several mathematics topics in every problem-solving phase in conceptual and procedural knowledge. This study aims to develop a mathematical problem-solving test as a mathematics DSPK test and investigate the content and construct validity. The product of test development is a 30-multiple-choice-item test in six mathematics topics. Every topic underlined all problem-solving phases in conceptual and procedural knowledge in a science or individual context. There were six experts performed the content validity sheets which analyzed using the content validity index (CVI) and intraclass correlation coefficient (ICC). The construct validity was examined using the Rasch model from 175 data of 7th grader students in Indonesia (Mage=12.66, SD=55). The result of content validity revealed overall items were valid (CVI≥83) and reliable (α=863; rxx=513). The construct of all items indicated fit (90≤weigheted MNSQ≤1.16) and were reliable (α=74) with various levels of difficulties and six low discrimination items. The recommendation for improvement is emphasized in language aspects. The absence of knowledge of facts could be an improvement for further study.

Corresponding Author:

Ijtihadi Kamilia Amalina
Doctoral School of Education, University of Szeged
6722 Szeged, Petőfi S. sgt.32-34, Hungary
Email: ijtihadi.kamilia.amalina@edu.u-szeged.hu

1. INTRODUCTION

Problem-solving is the goal of education and part of the science and mathematics curriculum [1]. The higher-order thinking and problem-solving model in the monodisciplinary area were developed in the 1960s' with the assumption by learning problem-solving, a student could transfer to any situation in daily life [1], [2]. Problem-solving is a complex set of cognitive, behavioral, and attitude components seen as a situational and context bond process that depends on the deep structures of knowledge and experiences [2], [3].

The development of problem-solving research is based on the theory that people encounter complex problems in their life, which require transdisciplinary [4], [5]. Moreover, the development of technology and economic globalization requires students to be good at transdisciplinary problem-solving in science, technology, engineering, and mathematics [5]. Hence, science, technology, engineering, and mathematics (STEM) problem-solving is infused into education. Since STEM problem-solving involves complex components in several disciplines, STEM problem-solving demands mastering domain-specific prior knowledge (DSPK) [6].

Several cognitive factors influencing problem-solving are prior knowledge, knowledge-based, formal reasoning ability, long-term memory, working memory, neo-Piagetian, and metacognitive variables [1], [2]. The most crucial factor is domain-specific knowledge under the theory of prior knowledge [2], [6], [7]. Prior knowledge is a stock of information, skills, beliefs, and experiences located before the learning process [8]. Other prior knowledge terms are pre-storage, permanent storage knowledge, background knowledge, pre-existing knowledge, and domain-specific knowledge [8]. Domain-specific knowledge engages in organizing new knowledge, encoding and representing information, assimilating new material, selecting relevant information to be elaborated, recalling, and retrieving information [2], [8]–[10].

In STEM problem-solving, it is essential to investigate students' DSPK in mathematics because mathematics plays a main part in the concept and content applications. However, there is a lack of study regarding measuring DSPK or developing assessment DSPK tools in mathematics. Several researchers focused on physics and biology [11]–[14], chemistry [15]–[17], and health education [18]. They implemented multiple-choice test [11], [18], two-tiered instrument [17], open-ended essay, complex essay, and map concept task [12], [13]. According to these issues, developing a mathematics DSPK instrument is necessary.

DSPK positively impacts problem-solving skills through experimental and non-experimental design research. Studies in DSPK in teaching and learning for primary to undergraduate students revealed that activating DSPK affects success in complex and creative problem-solving as well as self-regulation [1], [3], [4]. Non-experimental studies were conducted to find the correlation and impact of DSPK in mathematics on problem-solving. DSPK in mathematics was proven to impact students' complex and creative problem-solving skills [6], [19]. Moreover, studies about mathematics DSPK, specifically declarative (conceptual) and procedural knowledge, correlated and affected mathematical problem-solving [7], [20]–[23]. As a matter of fact, mathematics DSPK accounted for 56% of the variance [7], and conceptual understanding explained 31% of the variance in problem-solving [21].

DSPK in mathematics could be declarative (conceptual), procedural, strategic, and situational knowledge [2], [4], [12], [20]. Mostly, the researchers conducted procedural and conceptual research by applying essays (open-ended and short answers) and multiple-choice tests [4], [20]–[23]. However, they only focused on applying a phase in a topic rather than on all problem-solving phases in several mathematics topics. Moreover, DSPK assessment tools should concern the knowledge researchers want to assess in their main test. Hence, they could correlate with each other and measure the related knowledge. Consequently, not all previous assessment tools in mathematic DSPK could be adapted. According to the importance of mathematics DSPK (both procedural and conceptual knowledge) in complex problem-solving, in this case, STEM problem-solving, and the need for assessment tools on it, hence we discuss some issues in this paper regarding i) How is the development of the mathematical problem-solving test to measure DSPK in STEM problem-solving?; ii) How is the content validity of the developed mathematical problem-solving test?; and iii) How is the construct validity of the developed mathematical problem-solving test?

2. THEORETICAL BACKGROUND

2.1. Domain-specific prior knowledge, mathematical problem-solving, and their assessment tools

Prior knowledge is categorized as content knowledge and metacognitive knowledge [24]. Content knowledge is knowledge about domain or subject-oriented knowledge that contains procedural and conceptual (declarative) knowledge. Metacognitive knowledge refers to what a person knows about themself. It is divided into task knowledge, strategic knowledge, and self-knowledge.

Procedural knowledge is knowledge about how and when to use the appropriate skills and procedures [4], [20], [21]. It is the knowledge that contains actions and manipulations that are valid within a domain [2]. Conceptual knowledge, known as declarative knowledge in some resources, is knowledge about what and why a procedure is appropriate for a task, including checking the reasonable of the procedure [21]. It is knowledge about concepts, facts, definitions, operations, relations, and principles [2], [20]. Declarative knowledge could be assessed using multiple-choice and short answers, procedural knowledge is measured mainly by performance tests and possible by using multiple-choice and short answers, and strategic knowledge is rarely to be measured directly [2].

Hailikari, Nevgi, and Ylänne [24] distinguished declarative and procedural knowledge based on the cognitive domain. Declarative knowledge is divided into the knowledge of fact (KF) and knowledge of meaning (KM). The lower level of declarative knowledge has a low level of abstraction with simple recognition (i.e., recognition, enumerating, recalling, and remembering). The higher level of declarative knowledge needs deeper and requires the ability to define the concept (i.e., defining, reproducing, and understanding the meaning of the concept). Procedural knowledge is divided into the integration of knowledge (IK) and the application of knowledge (AK). The low level of procedural knowledge requires an ability to see the interrelation between concepts and how different phenomena link, including classifying and comparing. The higher level involves applying knowledge, performing a problem-solving task, producing, and implementing.

The types of knowledge in [24] engage in heuristic mathematical problem-solving phases, namely understanding a plan, devising the plan, carrying out the plan, and looking back. KM emphasizes the understanding phase, while KF does not include it in the problem-solving phase due to low abstraction knowledge. IK focuses on understanding, planning, and evaluating phases, while AK is concerned with applying phase.

Studies regarding the positive effects of DSPK assessments in complex and creative problem-solving were conducted using multiple-choice and constructed item tests [4], [6], [19]. The multiple-choice test focused on the relationship between variable and action knowledge (declarative and procedural knowledge) concerning the decision-making test [4] and conceptual knowledge [6]. DSPK related to conceptual (declarative) and procedural knowledge impacted mathematical problem-solving measured by constructed and selected types of tests [7], [20], [21], [25]. Some studies focused on algebraic problem solving (including proportion) and the prior knowledge related to it (i.e., fraction and proportion) [25] and algebra [21]. The basic concept knowledge and problem representation were also assessed by using short reasoning answer test [21]. A similar test was applied by [25] and concluded that conceptual knowledge about fractions is the most correlated factor (r=45) influencing proportional problem-solving.

2.2. Misconceptions, difficulties, and error in mathematics

Error in mathematics is caused by misconception, carelessness, reading problems, and some other factors. A misconception is different from an error since a misconception is a lack of understanding of a mathematics concept or rule [26]. A misconception about mathematics basic concepts was still found. There were 33.9% of Indonesian students had difficulties ordering arithmetic calculations, operating the whole number, and understanding negative signs [27]. In addition, students have difficulties converting length, weight, and time because they lack experience in their daily life [26].

The one-step higher concept is a fraction that is important to master higher concepts in algebra and statistics. However, students encountered misconceptions related to the meaning of the fraction (e.g., the meaning of fraction as a part of a whole, the meaning of fraction as a quantity (a quotient relation between two numbers) rather than two separate whole numbers). Moreover, they had problems adding fractions with the whole number, converting a fraction into a different representation (e.g., decimal, percentage), and calculating decimals because of the place value [26]–[28]. The related topic with the fraction is proportion. It was found that students had difficulties in conceptual, procedural, and factual knowledge [29]. They could not understand the concept of direct and indirect proportion. Hence, they could not distinguish between them. Regarding geometry topics, students had problems with both conceptual and procedural. They could not differentiate between the formula of area or circumference and volume or surface area, served 3D figures as 2D figures, and had a misconception about units [30].

The most serious difficulty in mathematical problem-solving is understanding the problem. Researchers interviewed primary teachers and concluded that students had difficulties in: i) Understanding keywords appearing in the problem; ii) Figuring out what to assume and what information is necessary; and iii) Reading and motivation problems [31]. When students have difficulties in the first phase of problem-solving, they cannot go to the higher phases (i.e., applying and evaluating).

3. RESEARCH METHOD

3.1. Instrumentation

The test is a 30-multiple-choice curriculum-based knowledge test. It measures students' conceptual and procedural knowledge of mathematical problem-solving. The topics used are numbers (integer and fraction) & measurement, ratio & proportion, geometry, and statistics. Each topic consisted of understanding (conceptual knowledge), planning, applying, and evaluating (procedural knowledge) phases in a science or individual context. There are four options for each question. The total score is 30 in a 90-minutes test. They completed the test online and could not use any helping device (e.g., a calculator). At the end of the test, there are demographic questions and a space for adding comments related to the drawback of the instrument (time, number of items, media, and difficulty).

The tool for assessing content validity is an assessment sheet that experts will complete. It contains three aspects: content, construct, and language aspects. The content aspect includes the suitability of aims, curriculum, indicators, and students' level. The construct aspect refers to the construction of questions, options, and information. The language aspect measures both clarity and ambiguity. Experts have to rate with 4 scales (e.g., 1=not clear and 4=very clear). The qualitative assessment is also provided by giving suggestions on every item for improvement.

3.2. Procedure

The procedure for developing the mathematical problem-solving test consisted of three main phases. The first phase analyzes the curriculum, students' needs, and topics we want to use in the test. The second phase generates the test indicators. The third phase generates the questions and options based on the indicators.

After developing the instrument, content validation was performed by expert panels. The experts' panels rated the instrument by using the assessment sheet. The expert panels are recruited based on several criteria: i) A minimum of a master's degree in mathematics or mathematics education; ii) Educators; and iii) A minimum of three years of working experience. The qualitative assessment from experts in each item will be used to identify and improve items that were mathematically inaccurate, ambiguous, or item prompting student responses that did not indicate their understanding of conceptual and procedural knowledge in mathematical problem-solving. It was validated after revising a draft by administering it to secondary school students. The objective of this phase is to verify that the test had good construct validity and reliability.

3.3. Participants

Six experts were selected to assess content validity, a male and five females. Of those, two lecturers in mathematics education and four mathematics teachers (three graduated from mathematics education, and one from mathematics). They graduated from different universities and currently work in different institutes in urban areas. The participants for investigating construct validity were 175 7th grader students from public schools in East Java, Indonesia (Mage=12.66, SD=55). Of those, 85 male and 90 female with 98.35% identified as Javanese, 1.1% were Madurese, and 6% were Sundanese.

3.4. Data analysis

The content validity was analyzed using the content validity index (CVI) and intraclass correlation coefficient (ICC). The construct validity was examined by using Rasch analysis. It described the item and person's behavior, including the item fit model, discrimination, item difficulties, and behavior of options. The analysis was performed with SPSS 25 and Conquest applications.

4. RESULTS AND DISCUSSION

4.1. Developing mathematical problem-solving test

STEM problem-solving requires complex skills, knowledge, cognitive components, and disciplines [2], [3]. Hence, it is needed to assess mathematics domain-specific knowledge in STEM problem-solving [8]. The conceptual and procedural knowledge in DSPK is important to be assessed because it influences problem-solving skills [7], [20]–[23]. Moreover, most Indonesian students had difficulties with basic concepts of mathematics. Hence, based on needs analysis, it requires developing a test for assessing DSPK in conceptual and procedural knowledge. Conceptual knowledge is divided into the knowledge of fact (KF) and knowledge of meaning (KM), and procedural knowledge is categorized into the integration of knowledge (IK) and the application of knowledge (AK) [24]. Since knowledge of fact has a low level of abstraction, it is not included in the development of the mathematical problem-solving test.

A multiple-choice test is the most appropriate type of test to measure DSPK [2], [4], [6]. Hence, we develop a multiple-choice test with four options. The options for every item have distractions. The distractions are the consequence of choosing the wrong answer in the previous number (or phase), some common misconceptions, common errors, and difficulties that students encounter. The misconceptions, difficulties, and common errors are: i) Interpreting text [31]; ii) Understanding the meaning of fractions [26]; iii) Converting a fraction into decimal and percentage [27]; iv) Ordering calculation/properties of number [27]; v) Understanding the decimal point/place value [26], [28]; vi) Figuring out what to assume, what is asked, what information from the problem is necessary to solve [31]; vii) Understanding conversion [26]; viii) Understanding and differentiating direct and inverse proportions [29]; ix) Differentiating between area and perimeter [30].

The test is targeted at junior secondary students (7th-9th graders), but we validated it on 7th graders only. The topics used are the topics applied in the main test and intersection topics for all graders based on the Indonesian curriculum and literature review. The topics are number (integer and fraction (n=9)) & measurement (n=4), ratio & proportion (n=8), geometry (n=4), and statistics (n=5). Each topic consisted of understanding, planning, applying, and evaluating phases in a science or individual context. KM measures students understanding phase of mathematics basic concepts (e.g., the meaning of fraction concepts). IK targets interrelation among concepts in understanding, planning, and evaluating phases (e.g., finding the correct statement regarding their problem-solving process). AK focuses on the application of knowledge in applying phase (e.g., calculating the reduction of CO_2 emission). Figure 1 explains the example of the test.

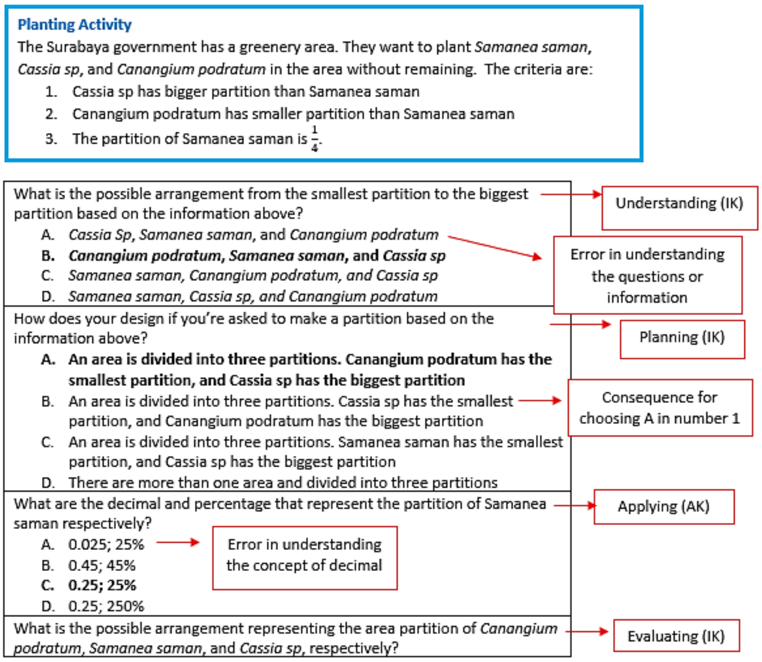


Figure 1. Example of the test in number topic

The indicators are from the Indonesian curriculum, mathematical problem-solving framework, and are related to the mathematics concept used in the main test. Table 1 explains the outline of the item for the mathematical problem-solving test. The total score is 30 in a 90-minutes test. The total time for conducting the test, answering demographic questions, and adding comments related to the developed test is 120 minutes.

Table 1. The outline of the item for the mathematical problem-solving test.

Basic competence	Item
Explain the concept of fractions and decide the order of integers and fraction	3
Solve problems related to ordering integers and fraction	1, 2, 5
Explain and calculate integers and fractions (and an approximation) by applying their properties	4, 8, 12
Solve problems related to integer calculation and a fraction (as well as an approximation) by applying integer properties	6, 7, 9, 10, 11, 13
Explain the concept of the ratio	18
Differentiate the concept of direct and inverse proportions by (or from) table, graph, and equation	19
Solve problems related to direct and inverse proportions by using a table, graph, or equation	14, 15, 16, 17, 20, 21
Solve contextual problems related to the area and perimeter/ circumference of the 2D figure	22, 23, 24, 25
Represent and solve a problem related to distribution, mean, median, modus, and the spread of data for making decisions or prediction	26, 27, 28, 29, 30

4.2. Content validity of the mathematical problem-solving test

There were two items (items 10 and 18) with a CVI of 83, an item with a CVI of 67 (item 11), and the rest with a perfect CVI value. Items 10 and 18 had a problem in the content aspect with a CVI of 83. These items were considered as a high level of difficulty. Item 11 had a problem in both content and language aspects with a CVI of 67, and it was categorized as an inadequate CVI [32]. Experts argued that item 11 had a high level of difficulty and ambiguity. Since item 11 was categorized as inadequate, it needs to be revised. Experts suggested providing more needed information and supplying new wording because of the high item difficulty.

The experts presented several suggestions mostly related to language aspects, including: i) Adjusting students' language ability; ii) Revising the ambiguous question; and iii) Supplying new wording to clarify the question. The ambiguous questions make readers misunderstand, leading them to different or double answers. The item that lacks information could produce a double answer as well. For example, an item requires students to find the length and width of a rectangle with an area of 20 cm^2, respectively. The options are “4 and 5” or “10 and 2”. The researcher wants students to choose “10 and 2”; hence, it is needed to add “l>w” to clarify.

Regarding the content, they met an agreement that the curriculum, indicator, and topics were appropriate for junior secondary students. However, several items were predicted to have a high level of difficulty. The given recommendation was to reduce the constraints or add valuable information. The evaluation phase items require students to decide on a true statement based on the given information. It is a high-difficulty item and requires students to reading comprehension. Therefore, re-phrase and supplying useful information were necessary. It was also recommended to add constraints in the low-level difficulty items (e.g., in the items related to numbers and measurements) and add relevant information in the high-level difficulty items (e.g., in the items related to geometry and statistics). The reliability of the mathematical problem-solving test was good, with a Cronbach Alpha of .863. Moreover, the intraclass correlation coefficient indicated moderate reliability with 51 [33].

4.3. Construct validity of the mathematical problem-solving test

The Rasch analysis was applied to investigate the construct validity with 60 iterations, 31 estimated parameters, and 6401.500 final deviance. All 30 items were categorized as fit, with MNSQ values ranging between 90 and 1.16. The value of weighted MNSQ is considered fit if it ranges between 6-1.4 [34]. The test could cover mostly all student abilities. However, several students with below-average abilities still could not solve the easiest item. Based on this theory of Rasch, it is recommended to add easier items or change the participants to a higher grade. Since it was a pilot study administered to 7th graders, the test targets 7th-9th grade students. Hence, it would be appropriate for them in the real test (the difficulty of the test will be appropriate for 8-9 grade students). The item difficulty level was distributed from the easiest level in item 1 (estimate value of -1.063) into the hardest level in item 30 (estimate value of 1.246). Item 1 had the easiest level because it only measures the simple understanding of given information with low constraints. Item 30 had the hardest level because it measures students' ability in evaluating phase, which needs higher-order thinking skills and complex calculations.

The item separation reliability indicates how well the item parameters are separated. In this case, the separation reliability was high with 941. The coefficient alpha value indicated the test was reliable at 74, and the EAP/PV reliability was 706. Figure 2 describes response model parameter estimates and Figure 3 represents a map of latent distributions and response model parameter estimates.

There were six items with low discrimination values, namely items 10 (discrimination of 14), 11 (discrimination of 08), 16 (discrimination of 08), 21 (discrimination of 16), 27 (discrimination of 17), and 29 (discrimination of 08). However, those items' MNSQ values were categorized as fit, and the options behaviors were acceptable. The addition result of a category (option) and correlation of those items represented acceptable value. The key answer had the highest value, except item 11. However, in items 6, 21, and 29, the probabilities of high-ability students clicking the different options from the key answers were higher, but the key answers still had the highest voter. After checking their previous related items, they clicked the right answers. Hence, their behavior in these items could be because of guessing or error. It could also occur because the options of these items consisted of a high level of distraction (seems similar option) that directed students to choose them.

VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
	item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1	1	-1.063	0.129	0.88	(0.79, 1.21)	-1.2	0.92	(0.90, 1.10)	-1.6
2	2	-0.857	0.128	0.96	(0.79, 1.21)	-0.4	0.97	(0.91, 1.09)	-0.7
3	3	-0.655	0.128	1.05	(0.79, 1.21)	0.5	1.05	(0.91, 1.09)	1.0
4	4	-0.832	0.128	0.97	(0.79, 1.21)	-0.3	0.99	(0.91, 1.09)	-0.3
5	5	-0.094	0.130	1.01	(0.79, 1.21)	0.1	1.02	(0.89, 1.11)	0.4
6	6	0.323	0.135	0.96	(0.79, 1.21)	-0.3	0.93	(0.85, 1.15)	-1.0
7	7	0.095	0.132	0.96	(0.79, 1.21)	-0.4	0.94	(0.88, 1.12)	-0.9
8	8	-0.555	0.128	0.88	(0.79, 1.21)	-1.1	0.90	(0.91, 1.09)	-2.3
9	9	0.538	0.138	0.86	(0.79, 1.21)	-1.4	0.90	(0.83, 1.17)	-1.1
10	10	0.040	0.131	1.13	(0.79, 1.21)	1.2	1.10	(0.88, 1.12)	1.7
11	11	0.670	0.140	1.19	(0.79, 1.21)	1.7	1.14	(0.81, 1.19)	1.4
12	12	0.444	0.136	0.89	(0.79, 1.21)	-1.0	0.91	(0.84, 1.16)	-1.1
13	13	0.264	0.134	1.12	(0.79, 1.21)	1.2	1.09	(0.86, 1.14)	1.3
14	14	-0.605	0.128	1.05	(0.79, 1.21)	0.5	1.03	(0.91, 1.09)	0.7
15	15	-0.680	0.128	0.98	(0.79, 1.21)	-0.2	0.98	(0.91, 1.09)	-0.5
16	16	0.382	0.135	1.22	(0.79, 1.21)	2.0	1.15	(0.85, 1.15)	1.9
17	17	-0.580	0.128	1.00	(0.79, 1.21)	0.0	1.02	(0.91, 1.09)	0.4
18	18	-0.832	0.128	0.97	(0.79, 1.21)	-0.3	0.97	(0.91, 1.09)	-0.6
19	19	0.635	0.140	0.90	(0.79, 1.21)	-0.9	0.94	(0.82, 1.18)	-0.6
20	20	-0.404	0.128	0.89	(0.79, 1.21)	-1.0	0.91	(0.91, 1.09)	-2.1
21	21	0.263	0.134	1.16	(0.79, 1.21)	1.4	1.10	(0.86, 1.14)	1.4
22	22	0.121	0.132	0.95	(0.79, 1.21)	-0.4	0.96	(0.87, 1.13)	-0.6
23	23	0.066	0.131	1.03	(0.79, 1.21)	0.4	1.00	(0.88, 1.12)	0.1
24	24	-0.122	0.130	1.12	(0.79, 1.21)	1.1	1.07	(0.89, 1.11)	1.3
25	25	0.845	0.144	0.92	(0.79, 1.21)	-0.7	0.93	(0.79, 1.21)	-0.7
26	26	0.234	0.133	0.91	(0.79, 1.21)	-0.9	0.90	(0.86, 1.14)	-1.4
27	27	0.066	0.131	1.15	(0.79, 1.21)	1.4	1.10	(0.88, 1.12)	1.6
28	28	0.635	0.140	0.94	(0.79, 1.21)	-0.5	1.00	(0.82, 1.18)	-0.0
29	29	0.412	0.136	1.22	(0.79, 1.21)	1.9	1.16	(0.85, 1.15)	2.0
30	30	1.246*	0.714	0.96	(0.79, 1.21)	-0.4	1.01	(0.73, 1.27)	0.1

An asterisk next to a parameter estimate indicates that it is constrained
Separation Reliability = 0.941
Chi-square test of parameter equality = 484.323, df = 29, Sig Level = 0.000
^ Quick standard errors have been used

Figure 2. Response model parameter estimates

```
                                            Item
                                           X|
                                           X|
                                         XXX|30
                                           X|
 1                                      XXXX|
                                         XXX|
                                        XXXX|25
                                     XXXXXXX|
                                     XXXXXXX|11
                                  XXXXXXXXXX|19 28
                                   XXXXXXXXX|9
                                     XXXXXXX|12 16 29
                                  XXXXXXXXXX|6
                                  XXXXXXXXXX|13 21 26
                                     XXXXXXX|22
                              XXXXXXXXXXXXXX|7 10 23 27
 0                        XXXXXXXXXXXXXXXXXX|
                       XXXXXXXXXXXXXXXXXXXXX|5 24
                         XXXXXXXXXXXXXXXXXXX|
                 XXXXXXXXXXXXXXXXXXXXXXXXXXX|
                XXXXXXXXXXXXXXXXXXXXXXXXXXXX|20
          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|
           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|8 17
                          XXXXXXXXXXXXXXXXXX|3 14
          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|15
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|4 18
         XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|2
-1  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|1
                     XXXXXXXXXXXXXXXXXXXXXXX|
                                XXXXXXXXXXXX|
                            XXXXXXXXXXXXXXXX|
                        XXXXXXXXXXXXXXXXXXXX|
                                     XXXXXXX|
                                       XXXXX|
                                      XXXXXX|
                                      XXXXXX|
                                        XXXX|
                                         XXX|
-2                                      XXXX|
                                           X|
                                           X|
                                          XX|
                                            |
                                           X|
Each "X" represents 0.3 cases
```

Figure 3. Map of latent distributions and response model parameter estimates

The behavior of distractors in item 11 was indicated as different from the theory. Most students clicked the C option (25.71%) rather than the right option, which is B (24%). The probability of students with a high ability level choosing option C was higher than choosing option B as shown in Figure 4. It was because of the language barrier. In item 11, the words "vertical" and "horizontal" made students encounter difficulty. Since the MNSQ value of item 11 was acceptable, we decided to revise rather than delete it [34].

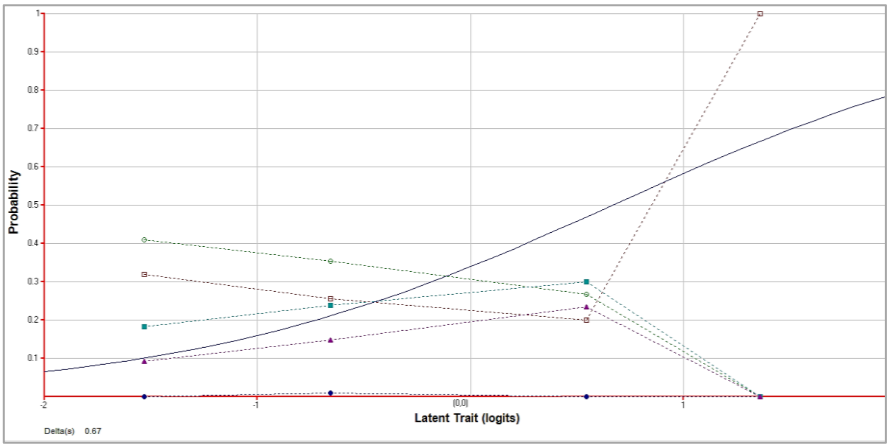


Figure 4. The characteristic curve of item 11 by category

Results of students' evaluation of the instrument revealed that 36% of students argued that they had a problem with the time given to solve the problem. We considered the time for solving the test to be from the Indonesian national examination, which is three minutes for every multiple-choice item. Regarding the test difficulty, 68% of students stated that the test was difficult. The number of items might influence the test difficulty. There were 40% of students expressed a problem with the high number of items. Since the test was multiple-choice, the probability of guessing is wide. A test with many items is required to avoid this issue. Moreover, some items and options need reading comprehension. There were 33% of students indicated that they had a problem with language understanding. Some items had been revised regarding language barriers. Regarding the media used, because the test is computer-based, there is a possibility for technical problems 6.3% of students encountered a technical problem during the test.

The test covered several contents and required higher-order thinking. Hence, the absence of knowledge of a fact becomes a drawback. Moreover, the number of items that measure knowledge of meaning, integration of meaning, integration of knowledge, and application of knowledge was not balanced. However, every phase of problem-solving in every content had occurred. The test was a curriculum-oriented test targeted at 7th-9th graders students. Hence it only could measure Indonesian junior secondary school students. In addition, the construct validity should be analyzed from sample grades 7th-9th, which was missing in this study.

5. CONCLUSION

The development of the problem-solving test serves as a DSPK test. It covers the limitation of previous studies by providing conceptual and procedural knowledge and every phase of problem-solving. It also provides a single assessment tool to measure students' DSPK in several mathematics contents. The product was a 30-item of multiple-choice test on six mathematics topics covered in a scientific or individual context.

The validation is a part of instrument development, including content and constructs validities. The mathematical problem-solving test had a good content validity index and reliability. Moreover, the construct validity indicated all items were fit. However, several items needed consideration and revision regarding the language barrier. The final product of the mathematical problem-solving test could be administered on a large scale to measure Indonesian junior secondary students' mathematical DSPK in complex problem-solving.

REFERENCES

[1] K. Serafino and T. Cicchelli, "Cognitive theories, prior knowledge, and anchored instruction on mathematical problem solving and transfer," *Education and Urban Society*, vol. 36, no. 1, pp. 79–93, Nov. 2003, doi: 10.1177/0013124503257016.

[2] J. J. Solaz-Portolés and V. S. López, "Cognitive variables in science problem solving: A review of research," *Journal of Physics Teacher Education Online*, vol. 4, no. 2, pp. 25–32, 2007.

[3] B. Rittle-Johnson, J. R. Star, and K. Durkin, "The Importance of Prior Knowledge When Comparing Examples: Influences on Conceptual and Procedural Knowledge of Equation Solving," *Journal of Educational Psychology*, vol. 101, no. 4, pp. 836–852, 2009, doi: 10.1037/a0016026.

[4] H. M. Süß and A. Kretzschmar, "Impact of cognitive abilities and prior knowledge on complex problem solving performance - Empirical results and a plea for ecologically valid microworlds," *Frontiers in Psychology*, vol. 9, no. MAY, pp. 1–22, 2018, doi: 10.3389/fpsyg.2018.00626.

[5] T. R. Kelley and J. G. Knowles, "A conceptual framework for integrated STEM education," *International Journal of STEM Education*, vol. 3, no. 1, 2016, doi: 10.1186/s40594-016-0046-z.

[6] C. Y. Chang, "Does problem solving = prior knowledge + reasoning skills in earth science? An exploratory study," *Research in Science Education*, vol. 40, no. 2, pp. 103–116, 2010, doi: 10.1007/s11165-008-9102-0.

[7] C. Oksuz, "Association of domain-specific knowledge and analytical ability with insight problem solving in mathematics," *International Journal of Pedagogies and Learning*, vol. 5, no. 1, pp. 138–153, Jan. 2009, doi: 10.5172/ijpl.5.1.138.

[8] T. Hailikari, *Assessing university students' prior knowledge. Implications for theory and practice*. Helsinki University Print, 2009.

[9] N. M. Crooks and M. W. Alibali, "Noticing relevant problem features: Activating prior knowledge affects problem solving by guiding encoding," *Frontiers in Psychology*, vol. 4, pp. 1–10, 2013, doi: 10.3389/fpsyg.2013.00884.

[10] D. C. Geary, D. B. Berch, R. J. Ochsendorf, and K. M. Koepke, *Acquisition of complex arithmetic skills and higher-order mathematics concepts*. Elsevier Inc., 2017.

[11] J. Wang, A. Stebbins, and R. E. Ferdig, "Examining the effects of students' self-efficacy and prior knowledge on learning and visual behavior in a physics game," *Computers and Education*, vol. 178, p. 104405, 2022, doi: 10.1016/j.compedu.2021.104405.

[12] T. Binder, A. Sandmann, B. Sures, G. Friege, H. Theyssen, and P. Schmiemann, "Assessing prior knowledge types as predictors of academic achievement in the introductory phase of biology and physics study programmes using logistic regression," *International Journal of STEM Education*, vol. 6, no. 1, 2019, doi: 10.1186/s40594-019-0189-9.

[13] T. Binder, P. Schmiemann, and H. Theyssen, "Knowledge acquisition of biology and physics university students—the role of prior knowledge," *Education Sciences*, vol. 9, no. 4, pp. 1–18, 2019, doi: 10.3390/educsci9040281.

[14] S. Kantahan, P. Junpeng, S. Punturat, K. N. Tang, P. Gochyyev, and M. Wilson, "Designing and verifying a tool for diagnosing scientific misconceptions in genetics topic," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 9, no. 3, pp. 564–571, 2020, doi: 10.11591/ijere.v9i3.20544.

[15] O. Gulacar, A. Milkey, and S. Mclane, "Exploring the effect of prior knowledge and gender on undergraduate students' knowledge structures in chemistry," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 15, no. 8, 2019, doi: 10.29333/ejmste/106231.

[16] I. G. Ionas, D. Cernusca, and H. L. Collier, "Prior knowledge influence on self-explanation effectiveness when solving problems : An exploratory study in science learning," *International Journal of Teaching and Learning in Higher Education*, vol. 24, no. 3, pp. 349–358, 2012.

[17] A. Versprille *et al.*, "Assessing student knowledge of chemistry and climate science concepts associated with climate change: Resources to inform Teaching and learning," *Journal of Chemical Education*, vol. 94, no. 4, pp. 407–417, 2017, doi: 10.1021/acs.jchemed.6b00759.

[18] M. E. Beier and P. L. Ackerman, "Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real-world learning environment," *Psychology and Aging*, vol. 20, no. 2, pp. 341–355, 2005, doi: 10.1037/0882-7974.20.2.341.

[19] G. Molnár, S. Greiff, and B. Csapó, "Inductive reasoning, domain specific and complex problem solving: Relations and development," *Thinking Skills and Creativity*, vol. 9, pp. 35–45, 2013, doi: 10.1016/j.tsc.2013.03.002.

[20] M. A. Al-Mutawah, R. Thomas, A. Eid, E. Y. Mahmoud, and M. J. Fateel, "Conceptual understanding, procedural knowledge and problem-solving skills in mathematics: High school graduates work analysis and standpoints," *International Journal of Education and Practice*, vol. 7, no. 3, pp. 258–273, 2019, doi: 10.18488/journal.61.2019.73.258.273.

[21] J. L. Booth and J. L. Davenport, "The role of problem representation and feature knowledge in algebraic equation-solving," *Journal of Mathematical Behavior*, vol. 32, no. 3, pp. 415–423, 2013, doi: 10.1016/j.jmathb.2013.04.003.

[22] K. H. Khng and K. Lee, "Inhibiting interference from prior knowledge: Arithmetic intrusions in algebra word problem solving," *Learning and Individual Differences*, vol. 19, no. 2, pp. 262–268, 2009, doi: 10.1016/j.lindif.2009.01.004.

[23] C. Rivella, C. Cornoldi, S. Caviola, and D. Giofrè, "Learning a new geometric concept: The role of working memory and of domain-specific abilities," *British Journal of Educational Psychology*, vol. 91, no. 4, pp. 1537–1554, 2021, doi: 10.1111/bjep.12434.

[24] T. Hailikari, A. Nevgi, and S. Lindblom-Ylänne, "Exploring alternative ways of assessing prior knowledge, Its components and their Relation to student achievement: A mathematics based case study," *Studies in Educational Evaluation*, vol. 33, no. 3–4, pp. 320–337, 2007, doi: 10.1016/j.stueduc.2007.07.007.

[25] A. K. Jitendra, A. E. Lein, J. R. Star, and D. N. Dupuis, "The contribution of domain-specific knowledge in predicting students' proportional word problem solving performance," *Educational Research and Evaluation*, vol. 19, no. 8, pp. 700-716, 2013, doi: 10.1080/13803611.2013.845107.

[26] R. G. Mohyuddin and U. Khalil, "Misconceptions of students in learning mathematics at primary level," *Bulletin of Education and Research*, vol. 38, no. 1, pp. 133–162, 2016.

[27] R. F. Jabal and R. Rosjanuardi, "Identifying the secondary school students' misconceptions about number," *Journal of Physics: Conference Series*, vol. 1157, no. 4, 2019, doi: 10.1088/1742-6596/1157/4/042052.

[28] H. R. Hj Sarwadi and M. Shahrill, "Understanding students' mathematical errors and misconceptions: The case of year 11 repeating students," *Mathematics Education Trends and Research*, vol. 2014, pp. 1–10, 2014, doi: 10.5899/2014/metr-00051.

[29] M. Oktaviani, "Analysis of students' error in doing mathematics problem on proportion," In *2nd Asian Education Symposium - AES*, 2018, pp. 172–177, doi: 10.5220/0007300601720177.

[30] G. Tan Sisman and M. Aksu, "A study on sixth grade students' misconceptions and errors in spatial measurement: Length, area, and volume," *International Journal of Science and Mathematics Education*, vol. 14, no. 7, pp. 1293–1319, 2016, doi: 10.1007/s10763-015-9642-5.

[31] P. Phonapichat, S. Wongwanich, and S. Sujiva, "An analysis of elementary school students' difficulties in mathematical problem solving," *Procedia - Social and Behavioral Sciences*, vol. 116, pp. 3169–3174, 2014, doi: 10.1016/j.sbspro.2014.01.728.

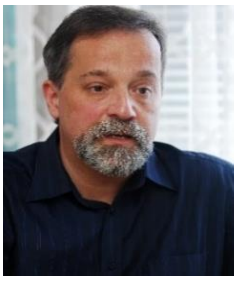
[32] K. Hyrkäs, K. Appelqvist-Schmidlechner, and L. Oksa, "Validating an instrument for clinical supervision using an expert panel," *International Journal of Nursing Studies*, vol. 40, no. 6, pp. 619–625, 2003, doi: 10.1016/S0020-7489(03)00036-1.

[33] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, 2016, doi: 10.1016/j.jcm.2016.02.012.

[34] T. G. Bond and C. M. Fox, *Applying the rasch model: Fundamental measurement in the human sciences*. New York: Routledge, 2015.

BIOGRAPHIES OF AUTHORS

Ijtihadi Kamilia Amalina is a Ph.D. candidate, Doctoral School of Education, University of Szeged, Hungary. She received her bachelor and master degrees in mathematics education from Universitas Negeri Surabaya, Indonesia. Her research focuses on assessment, mathematics education, STEM, social-emotional, and problem-solving. She can be contacted at email: ijtihadi.kamilia.amalina@edu.u-szeged.hu.



Tibor Vidákovich is a professor at the Institute of Education, University of Szeged, Hungary. He obtained his PhD in education at the Hungarian Academy of Sciences in 1991, and he was appointed a university professor in 2010. His research interests include educational assessment, diagnostic testing, development of tests and test banks; and the development and enhancement of deductive reasoning, mathematical abilities and mathematical competence. He can be contacted email: t.vidakovich@edpsy.u-szeged.hu.