



# Does a self-report questionnaire predict strategy use in mathematical problem solving among elementary school children? Importance of question format depending on the grade

Shiho Kashihara<sup>1,3</sup> · Tatsushi Fukaya<sup>2</sup>

Received: 9 August 2022 / Revised: 16 November 2022 / Accepted: 30 November 2022 /

Published online: 22 December 2022

© The Author(s) 2022

## Abstract

Predicting the actual performance of strategy use with self-report questionnaires is difficult, especially among elementary school children. Nevertheless, due to the simplicity of self-report questionnaires, it is desirable to identify one that can predict children's performance of actual strategy use. This study investigated whether a self-report questionnaire on the frequency of using a diagram strategy in mathematical problem-solving can predict children's actual diagram use by manipulating the question type (i.e., free description, multiple-choice, and Likert scale). We also examine the question types that can better predict the actual strategy use in elementary school children. Fourth- to sixth-grade children were asked to complete both a questionnaire, which asked about their daily use of a diagram strategy through three question types and a test consisting of math word problems. We found that when children were asked to self-report their strategy using a Likert scale, they were predicted to use diagrams during the test regardless of grades. Furthermore, the older the children became, the more effective it was to ask them to self-report in a free description type. These results suggest that appropriate question types can make it possible to measure actual strategy-use behaviors through self-report measures, even for elementary school children.

**Keywords** Learning strategy · Self-report questionnaire · Strategy of diagram use · Elementary school children

---

<sup>1</sup>Shiho Kashihara is now at the Department of Cognitive Neuroscience, Cognitive Mechanisms Laboratories, Advanced Telecommunications Research Institute International (ATR), Keihanna Science City, Kyoto 619-0288, Japan.

---

✉ Tatsushi Fukaya  
fukaya@hiroshima-u.ac.jp

<sup>1</sup> Graduate School of Education, Hiroshima University, Hiroshima, Japan

<sup>2</sup> Graduate School of Humanities and Social Sciences, Hiroshima University, 1-1-1 Kagamiyama, 739-8524 Higashi-Hiroshima, Hiroshima, Japan

<sup>3</sup> Research Fellow of Japan Society for the Promotion of Science, Hiroshima, Japan

## Introduction

Several studies have examined the use of learning strategies, such as what strategies are used and which are effective for learning (e.g., Bjork et al., 2013; Dunlosky et al., 2013; Edwards et al., 2014; Tullis & Maddox, 2020). According to Derry and Murphy (1986), learning strategy is defined as “the collection of mental tactics employed by an individual in a particular learning situation to facilitate acquisition of knowledge or skill” (p. 2). A major research approach to learning strategies involves investigating their use with self-report questionnaires. Several self-report questionnaires are widely used, for instance, the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich & de Groot, 1990) and Learning And Study Strategies Inventory (LASSI; Weinstein et al., 1988), to assess the extent to which learners use strategies (e.g., Bråten & Olaussen, 1998; Cano, 2006; Duncan & McKeachie, 2005; Sungur, 2007). Measuring strategy use with self-report questionnaires can be easily conducted for many participants at once and is more effective to implement than the behavioral observation of actual strategy use (e.g., Berger & Karabenick, 2016; Jacobse & Harskamp, 2012; Schellings & Van Hout-Wolters, 2011).

Regarding self-report instruments, especially on general learning strategies, methodological drawbacks such as varying reference points of each respondent or social desirability of answers have been indicated (Richardson, 2004; Schellings et al., 2013). A challenge is that answers to self-report questionnaires are sometimes poorly related to the extent of learners' actual strategy use (e.g., Desoete, 2008; Kikas & Jögi, 2016; Lehmann et al., 2022; McNamara, 2011; Saraç & Karakelle, 2012; Schellings & Van Hout-Wolters, 2011). For example, Veenman (2011) noted that offline indices requiring retrospective judgments have low validity in clarifying strategy use. Further, other researchers found no significant correlations between the self-report questionnaire and either the task performance of mathematical problem-solving or online indices, including the think-aloud method and observation of traces of strategy use during problem-solving (Jacobse & Harskamp, 2012).

While offline instruments may have limited validity, self-report measurements can still be used to evaluate actual strategy use under certain conditions. Researchers have explored the types and situations of self-report measures that can adequately measure actual behavior, for example, using a task-specific questionnaire that is directly based on a taxonomy used for coding online measurement protocols (Schellings, 2011) or asking participants to choose an appropriate option from multiple-choice questions that described a situation of strategy use (Cromley & Azevedo, 2011). Given the advantage of self-report measurements for simplicity and convenience, it is important to clarify when self-report measures are useful in predicting actual strategy use behavior. However, there is still uncertainty regarding whether and what kinds of self-report questionnaires are efficient in measuring actual strategy use.

There is a great need to identify which questionnaires are relevant to actual strategy use, especially for elementary school children. Questionnaires with low implementation cost and high convenience are particularly effective for younger age groups; however, the validity of self-reported metacognition is lower among elementary school children than among junior high school, high school, and university students (Craig et al., 2020). Craig et al. (2020) conducted a meta-analysis on the relationship between online control tasks and offline self-reported knowledge in measuring metacognition. The authors showed that both correlation and heterogeneity (i.e., the magnitude of variation in results across studies;  $I^2$ ) were smaller

for primary students ( $r=.08$  (95% CI = [-.05, .21]),  $I^2=10\%$ ) than for secondary ( $r=.33$  [-.002, .65],  $I^2=52\%$ ) and university students ( $r=.29$  [.18, .40],  $I^2=59\%$ ). That is, there was little association between the online and offline indices consistently across studies, especially among elementary school children. This may suggest age-related differences in metacognitive skills and that elementary school children have underdeveloped metacognitive skills due to inexperience and insufficient expertise; therefore, they may not know how to apply the knowledge to a task or be fully aware of their strategy use (Craig et al., 2020). Thus, it is necessary to reveal a valid self-report measure that can adequately predict strategy use, especially for elementary school children developing metacognition.

This study aimed to clarify whether and how a self-report questionnaire can predict actual strategy use in fourth- to sixth-grade elementary school children using several question formats. To enhance the validity of the self-report questionnaire, we focused on a specific strategy rather than a general one and targeted fourth- to sixth-graders among elementary school children.

First, we focused on the specificity of the target strategy. When examining a valid self-report measure that is a good predictor of actual strategy use behavior, the task specificity of the target must be considered (Craig et al., 2020; Schellings, 2011). Samuelstuen and Bråten (2007) reported that a self-report questionnaire that measures general strategy use has low validity compared to the measurement of specific strategies in a particular task context in 15-year-old students. Additionally, the task-specific questionnaire for measuring metacognitive skills in mathematics problem-solving better correlated with the performance from the think-aloud paradigm than the general questionnaire (Veenman & van Cleef, 2019). Although these findings are based on junior high school and high school students, they have suggested that the association between self-report questionnaires and the performance of actual strategy use is affected by the context of the target strategy. It is more likely that assessing the degree of strategy use for task-specific situations is more valid than measuring general strategy use, that is, individual tendencies of strategy use regardless of a situation or context. Since elementary school children have less ability to generalize a situation than older students, it is desirable to focus on a specific strategy in a specific learning context.

Therefore, this study asked elementary students to report their strategy use, especially diagram use in mathematical problem-solving (e.g., Hembree, 1992; Schoenfeld, 1985; Uesaka et al., 2007). Diagram use is one of the most effective and popular strategies to promote one's performance in solving mathematical problems (e.g., Hembree, 1992; Lowrie, 2020). Uesaka et al. (2007) examined the relationship between self-reported daily use of diagrams and actual strategy use when solving mathematical problems in junior high school students. They reported a moderate positive correlation between the answer to the self-report questionnaire and the actual behavior of strategy use. However, the participants of Uesaka et al.'s (2007) study were 13-to 15-year-olds; therefore, it is still unclear whether such relationships between self-report answers and the performance of diagram use are also observed in elementary school children, who are younger, and have less validity in measuring metacognition via self-report than do junior high school and older students.

Second, we targeted fourth- to sixth-graders. Studies suggest that elementary school children's metacognitive skills are not sufficiently developed (Craig et al., 2020; Dermitzaki, 2005; Haberkorn et al., 2014; Metcalfe & Finn, 2013). However, elementary school students are of a wide range of developmental ages from six to twelve years old (in the case of Japan), and their metacognitive abilities are different. Indeed, previous studies have suggested that

an age-related improvement in metacognitive control ability appears in upper elementary school children at approximately 10 years old (e.g., Bayard et al., 2021; Dufresne & Kobasigawa, 1989; Roebers et al., 2014). Dufresne and Kobasigawa (1989) investigated the metacognitive control abilities of first-, third-, fifth-, and seventh-grade children concerning allocating learning time and found that fifth- and seventh-grade students could allocate their learning time more efficiently on the level of difficulty of the study. Roebers et al. (2014) showed other examples of developmental differences in children's metacognition; they found a developmental improvement in metacognitive monitoring and control, and closer associations between metacognitive process and test performance in eleven-year-olds than in nine-year-olds. Thus, considering the development of children's metacognition, accurately judging their strategy use is still a developing skill even in the middle to upper elementary grades, and as children get older, they are likely to be more accurate in reporting their strategy use behavior. Therefore, this study specifically focused on the upper grade of elementary school children, who are nine- to twelve-year-olds.

Additionally, as a new perspective, we focused on the impact of question formats. Most self-report questionnaires use the Likert scale, which confirms how often one uses a certain strategy in a four- or five-point format (e.g., Pintrich & de Groot, 1990; Weinstein et al., 1988). However, other types of questionnaires besides the Likert format exist, for example, those that ask participants to answer freely about their strategy use (e.g., "What kind of strategies do you use when you are studying?"; Karpicke et al., 2009; Zimmerman & Martinez-Pons, 1988) or those that ask participants to choose from several possible answers regarding whether they would use a certain strategy in a particular scenario (e.g., the following scenario in Karpicke et al., 2009: "Imagine you are reading a textbook chapter for an upcoming exam. After you have read the chapter one time, would you rather ..."). The effect of questionnaire formats on tasks is a key topic examined in the past (e.g., Jönsson et al., 2017; Rodriguez, 2003). However, despite the suggested effects of different question formats, how the relationship with objective indices varies with the various question formats of self-report questionnaires has not been examined.

When asking about strategy use, the degree of difficulty may differ between the question formats: the multiple-choice format, wherein participants select the strategy they have used from a list of options; a Likert scale format, wherein participants choose the intensity of the frequency of use of each strategy; and the open-ended free-description format, wherein participants freely report the recalled strategies they use. Put differently, similar to how a recognition task is easier than a recall task for its format (e.g., Postman et al., 1975), multiple-choice and Likert formats are expected to be easier to answer than a free-description format questionnaire since there is no need to generate the choice options oneself and instead use various cognitive strategies (such as inference). Another possibility is that elementary students are not yet good at meta-thinking and therefore cannot answer the free-description format well. Thus, it is necessary to compare multiple question formats of self-report questionnaires on strategy use to determine the best format that predicts actual diagram use behavior. Especially for elementary school children in the process of metacognitive development, their responses may be greatly influenced by the question format. Therefore, it is possible to identify more efficient self-report indices by examining more appropriate question formats among self-report questionnaires. In this study, we prepared three types of self-report questionnaires: free-description, multiple-choice, and Likert. Sub-

sequently, we examined which of the three question formats better predicted actual diagram use in mathematical problem-solving.

Overall, this study aimed to examine an offline questionnaire measurement to predict actual performance of diagram use in mathematical problem-solving for fourth- to sixth-grade elementary school children, who have reported weak relationships between online and offline measurements. We focused on the diagram use strategy, a highly specific strategy for mathematics, to make it easier to examine relationships between online and offline indices in elementary school children. Additionally, we compared three formats of self-report questionnaires to examine the best predictor of actual strategy use. Alongside examining correlations between actual diagram use during word problem-solving and offline self-report measures, we also examined how each question format uniquely predicted actual strategy use in an exploratory analysis.

Since previous studies using highly specific strategies (i.e., diagram use) for examining the relationship between actual online strategy use and offline self-reported measurements have reported positive correlations between online and offline measures (Uesaka et al., 2007), we hypothesized that the two would be positively correlated even in elementary school children. Moreover, the correlation strength was expected to vary depending on the question format: the Likert format, which has the lowest abstraction or freedom of question, was expected to have a stronger association between online and offline indices. In contrast, the free-description format, requiring additional meta-monitoring ability of one's own activities, was expected to have a weaker association with the online index, especially for lower-grade children.

## Method

### Participants

A total of 496 third- to sixth-grade students in one of two public elementary schools in Tokyo, Japan (2–3 classes per grade) participated in the survey. The survey for the first school was conducted from April to May 201X for 179 fourth- to sixth-grade elementary school children (66 fourth-grade students, including 28 boys and 38 girls; 64 fifth-grade students, including 33 boys, 30 girls, and one of unknown gender; 49 sixth-grade students, including 28 boys, 18 girls, and three of unknown gender). The survey for the second school was conducted from December 201X to January 201X+1 for 317 third- to fifth-grade students (110 third-grade students, including 58 boys, 49 girls, and three of unknown gender; 104 fourth-grade students, including 53 boys and 51 girls; 103 fifth-grade students, including 43 boys, 53 girls, and seven of unknown gender).

Since the survey was conducted at the beginning of the school year (April) in the first school and the end of the school year (January) in the second school, the grades in the second school were treated in the same age group as those in the following year. For example, fourth-graders in the first and third-graders in the second school were treated as the same age group (fourth-graders). Therefore, there were 176 fourth graders (86 boys, 87 girls, and three of unknown gender); 168 fifth graders (86 boys, 81 girls, and one of unknown gender); and 152 sixth graders (71 boys, 71 girls, and 10 of unknown gender). Twenty-eight children were excluded for missing or unanswered responses. Thus, we analyzed data from 468 chil-

dren (171 fourth graders comprising 86 boys and 85 girls; 159 fifth graders comprising 81 boys and 78 girls; 138 sixth graders comprising 69 boys and 69 girls).

This study was approved by the Ethical Committee of the university to which the authors belonged. We obtained informed consent from the school principals of the children who participated in the survey. After the survey, we distributed examples of the ideas and solutions for the math problems used in the survey to the children. We also sent the survey feedback to teachers, including the correct answer rate of the mathematical problems, the percentage of children who used diagrams, and a summary of answers to the questionnaire, which can be used as material for future instruction.

## Inventory

For the convenience of conducting the survey, the inventory was organized into two booklets—questionnaire and test—by the implementation date.

**Questionnaire booklet.** It comprised seven questions, including one free-description question, one multiple-choice question, and five Likert-type questions. The items' order was fixed to ensure they did not affect each other: first, free-description, then multiple-choice, and finally, Likert question. The contents of the questionnaire booklet were the same for all grades.

In the free-description question, participants were asked, "How do you usually approach mathematical word problems when solving them? Please tell us the strategies you use to solve mathematical word problems effectively, being as specific and detailed as possible. Write not only one tip, but as many as you can recall." Subsequently, they were asked to fill in the blanks with free descriptions.

In the multiple-choice question, participants were presented with a specific situation and asked to choose one of the options that they would most likely opt for: "During a mathematical test, you are trying to solve a mathematical word problem, but you do not know how to solve it right now. In such a situation, what would you do? Choose one of the following options and circle the one that best describes your usual behavior in such situations. — (1) Give up solving it and move on to the next problem; (2) Do not give up and keep thinking in your mind; (3) Try adding, subtracting, multiplying, and dividing the numbers in the problem text by trial and error; (4) Draw a diagram; (5) Others (free description)."

In the Likert-type questions, participants were presented with five learning strategies for solving mathematical word problems, including one about diagram use ("To understand the meaning of the problem, I use diagrams and tables"), and asked, "How often do you use the following learning methods when you study mathematics? For each of the strategies, choose one frequency from *Never* (1) to *Often* (4) and circle the number." For each of the five strategies, participants answered the frequency of each strategy use through a four-point scale ("1: Never use," "2: Not use very often," "3: A little use," "4: Very much use").

**Test booklet.** The test booklet comprised four mathematical word problems based on the previous grade level for each grader. To make the use of diagrams more necessary, we prepared applied problems rather than basic problems that can be solved without using diagrams by modifying some Japanese commercial workbooks ("Syogaku 3/4/5-grade Hyojun Mondai-shu Sansu," published by Juken kenkyu-sya, in Japan). Moreover, to avoid the influence of the kinds of problems and increase the generalizability of the results, we prepared two versions of the test booklet, comprising four different mathematical word

problems for each grader. Subsequently, we asked teachers to give children test booklets based on their student numbers to ensure an equal number of versions 1 and 2 in each class. The test booklets included mathematical problems such as the following: “There are three children with two balloons in each hand. How many balloons are there in total?” (Version 1), and “There are many colored papers. When divided into 40 pieces each, you can make eight sheaves and have 20 pieces left. How many pieces of colored paper are there in total?” (Version 2) for fourth graders.

The test booklet was an A4 size sheet printed on both sides, with a cover page followed by a facing page of four word problems. Below each question, there was enough space to write notes, formulas, and diagrams. In all grades, participants were asked not to use erasers and write down whatever they thought by the instruction on the cover page.

## Procedure

The survey was conducted after obtaining survey permission from the principal of each school. We mailed the inventory booklets to each school and asked the homeroom teacher of each class to conduct the survey (e.g., distribute the booklets, instruct before answering the inventory, and collect the booklets) as per the researchers’ instructions. As noted, we asked the homeroom teachers that the two booklets of inventory (i.e., the test and questionnaire booklets) not be answered at the same time; rather, assessment of the test booklet should be conducted on different days during the same week after the administration of the questionnaire booklet, and the two versions of the test booklets (versions 1 and 2) should be distributed alternately following the order of student list in each class to avoid any particular bias in test booklet distribution.

## Data coding

The coding criteria for the correctness of the mathematical problems, use of diagrams for the test booklet, and answer to the free-description and multiple-choice questionnaires were prepared in advance by the authors in consultation with each other. After the scoring exercise, two raters independently coded the test and questionnaire according to the coding criteria. Data from the first school were scored by two authors, and data from the second school were scored by the first author and a university student majoring in elementary education. When the two raters did not agree on the item scores, the ratings were decided through a discussion.

**Coding of test performance.** For each participant, the number of correct answers among all four questions was independently evaluated by two raters. For each word problem, one point was given for a correct and zero for an incorrect answer. However, one point was still given even if the answer was incorrect, only when the correct answer was written in the memo, but children mistakenly copied it to the answer space. Additionally, 0.5 points were given if the formulation of an equation was correct but the answer was wrong (e.g., the decimal point was wrongly placed).

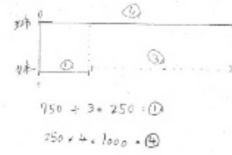
**Coding of actual diagram use.** For each participant, the appearance of the diagram use for each of the four-word problems was independently evaluated by two raters. For each question, one point was given if a diagram was used and zero if no diagrams were used. In this study, a diagram was defined as any coherent description showing at least two or more

(a) 問4 栗鼠3人で、このボールゲームを争います。お母さんが  $\frac{1}{2}$  点、お兄さんが  $\frac{1}{4}$  点、お姉さんが  $\frac{1}{4}$  点奪いました。ウーキーはどのくらいのことまでしよう。  
 [お母さんを含めるとお兄さん以下のスペースに書いてください。どのようなことを書いてもかまいませんが、誤字に気をつけてください。]



(表2)

(b) 問6 何と何の持っているお金をくらべて、誰は誰の4倍持っています。2人のお金の差は、750円です。終と始の持っているお金は、それぞれいくらですか。  
 [お母さんを含めるとお兄さん以下のスペースに書いてください。どのようなことを書いてもかまいませんが、誤字に気をつけてください。]



お母 = 250円  
 (表2) お兄 = 1000円

(c) 問3 高さ12cmの40個のけいねが並べられてあります。このけいねのある面から見て、2層目には1層より高さが増えました。32cmでした。はじめに何本のけいねがあるかを選んでください。  
 [お母さんを含めるとお兄さん以下のスペースに書いてください。どのようなことを書いてもかまいませんが、誤字に気をつけてください。]



(表2) 2m

(d) 問2 右で、積、高さのそれぞれを2倍にした立方体の体積は何倍になりますか。  
 [お母さんを含めるとお兄さん以下のスペースに書いてください。どのようなことを書いてもかまいませんが、誤字に気をつけてください。]

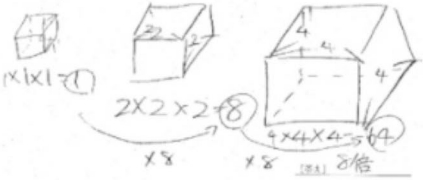


Fig. 1 Example of actual diagrams drawn by children. (a) shows a diagram drawn by a fourth grader, (b) shows a diagram drawn by a fifth grader, and (c) and (d) are diagrams drawn by a sixth grader

elements, such as digits and arrows or/and lines, to represent relations of variables independent of numerical formulas (cf., Uesaka et al., 2007). When the additional elements, such as lengths or lines, were written in the figures given in the word problems, those descriptions were also considered diagrams. Tables were also included in the diagrams. Additionally, even if the children deleted the diagrams, they were counted as diagrams if they could be identified. In contrast, following the definition of a diagram in this study, descriptions wherein elements or arrows were shown separately were not considered diagrams; simple conversions such as “1 cm → 100 mm” and formulaic expressions such as “percent triangle (*ku-mo-wa* in Japanese)” or “time-speed-distance triangle (*mi-ha-ji* in Japanese)” were not considered diagrams because they were just tools for simplifying calculations, not for understanding the situation in the word problem. An example of an actual diagram drawn by children is shown in Fig. 1.

After discussing the items on which the two raters did not agree, the percentages of correct answers and ratio of diagram use among all four-word problems were calculated for each participant. Additionally, the ratio of diagram use restricted to wrong answers was also calculated.

**Coding of the free-description scale.** A total of 14 categories of possible answers were generated by the discussion among authors in advance (e.g., “diagram uses”: descriptions referring to the use of diagrams, such as “draw diagrams,” “organizing the information in problems by making figures, tables, and graphs,” and so on; for all categories, see Table S1 in supplementary information). When a participant submitted two or more answers, each answer was divided by bullets or sentence meaning and was evaluated as a separate response. The ratio of agreement or reproducibility between the two raters was then calculated (i.e., kappa coefficient). The categories of items not classified in the agreement were



determined by discussion. Subsequently, the answers with or without reference to the category of diagram used were dummy coded as binary data (0/1) per participant.

**Coding of multiple-choice scale.** If participants chose “5: Others” and wrote down their strategies freely on the multiple-choice scale, the authors classified them into three categories generated by discussion among authors in advance (1: resolving, such as “After reaching the last problem, come back and solve it again”; 2: careful reading, such as “Reread the question carefully”; 3: others). Subsequently, the raters discussed categories for which no consensus was reached among the two raters, following the calculation of the agreement rate. This study’s focus was whether children chose option 4, “draw a diagram”; therefore, choosing that option or not was dummy coded as 0 (absence) or 1 (presence) for each participant.

**Likert scale.** Each of the five Likert-type questions was scored from one to four, with higher scores indicating more frequent use of the strategy and lower scores indicating a lower frequency of strategy use. Since this study focused on the response to the question, “To understand the context of the word problem, draw figures and tables to represent them,” the frequency of strategy use was scored on a four-point scale.

As a result of coding, the agreement rate (Kappa) was as follows: for coding of the test performance,  $\kappa_s=0.95\text{--}0.97$  ( $\kappa=0.95$  for the first school,  $\kappa=0.97$  for the second school); for coding of the actual diagram use,  $\kappa_s=0.93\text{--}0.96$  ( $\kappa=0.96$  for the first school,  $\kappa=0.93$  for the second school); for coding of the free-description question,  $\kappa_s=0.80\text{--}0.84$  ( $\kappa=0.84$  for the first school,  $\kappa=0.80$  for the second school); for coding of the free answers in the multiple-choice question,  $\kappa_s=0.84\text{--}0.90$  ( $\kappa=0.90$  for the first school,  $\kappa=0.84$  for the second school).

## Results

### Summary statistics

We analyzed the data using R 4.1.2 (R Core Team, 2021). For the data from 468 participants, after omitting missing data, we calculated for each grade level the average correct answer rate of the test and average rate of diagram use in the test, as well as the average referenced rate of diagrams in the answers to the free-description question, average selected rate of the option of diagram use in the multiple-choice question, and average frequency of diagram use in the Likert-type question (Table 1). When participants’ school and version of the test booklets were combined for each grade level, the average correct answer rate of the test was 51% ( $SD=30$ ) for fourth-graders, 47% ( $SD=36$ ) for fifth-graders, and 42% ( $SD=36$ ) for sixth-graders. Additionally, the average rate of actual diagram use in the test was 16% ( $SD=23$ ) for fourth-graders, 28% ( $SD=24$ ) for fifth-graders, and 37% ( $SD=33$ ) for sixth-graders. The overall correct answer rate was approximately 50% in all grades, confirming that the difficulty was set at a level hard to solve without diagram use. Additionally, for answers to the self-report questionnaire, the average rate of referring to diagram use (for the free description) was 10% ( $SD=30$ ) for fourth graders, 23% ( $SD=42$ ) for fifth graders, and 37% ( $SD=48$ ) for sixth graders. The average rate of choosing the strategy of diagram use (for the multiple-choice scale) was 13% ( $SD=34$ ) for fourth, 25% ( $SD=43$ ) for fifth, and 24% ( $SD=43$ ) for sixth graders. Finally, the average frequency of subjective reports of

**Table 1** Mean rates for each index in the test and self-report questionnaire

Grade	School	N	Test						Self-report questionnaire							
			Correct rate		Diagram rate		Free description		Multiple-choice		Likert (four-point)					
			Version 1 M	SD	Version 2 M	SD	Version 1 M	SD	Version 2 M	SD	M	SD	M	SD		
4th	A	66	0.65	0.28	0.55	0.29	0.17	0.23	0.12	0.17	0.14	0.35	0.17	0.38	2.95	0.90
	B	105	0.48	0.30	0.43	0.28	0.25	0.30	0.09	0.15	0.08	0.27	0.11	0.32	2.82	0.95
5th	A	57	0.56	0.34	0.62	0.38	0.29	0.18	0.32	0.28	0.19	0.40	0.23	0.42	3.11	0.86
	B	102	0.40	0.34	0.42	0.34	0.30	0.20	0.25	0.28	0.25	0.44	0.25	0.44	3.14	0.91
6th	A	45	0.55	0.36	0.30	0.27	0.22	0.27	0.30	0.27	0.13	0.34	0.07	0.25	3.00	0.83
	B	93	0.51	0.37	0.29	0.34	0.41	0.33	0.45	0.37	0.48	0.50	0.32	0.47	3.19	0.88

*Note.* For the test, both indices show the rate of the mean number of questions (correct or using a diagram). For the self-report questionnaire, free description, and multiple-choice are shown as the rate of the responses mentioning the use of diagrams out of the total number of responses. Furthermore, the Likert scale reflects the mean points

diagram use (for the four-point Likert scale) was 2.87 ( $SD=0.93$ ) for fourth graders, 3.13 ( $SD=0.89$ ) for fifth graders, and 3.13 ( $SD=0.86$ ) for sixth graders.

Since participants belonged to two schools and two versions of test booklets were used in the present data, we confirmed the effects of these two extraneous variables. We conducted a two-way multivariate analysis of variance per grade for all five test and questionnaire indices, with the independent variables of the participants' school and version of the test booklet, using Pillai's trace statistic. As a result, we found significant main effects of school affiliation (Pillai's trace=0.07,  $F(5,163)=2.56$ ,  $p=.03$  in fourth graders; Pillai's trace=0.07,  $F(5,151)=2.42$ ,  $p=.04$  in fifth graders; Pillai's trace=0.20,  $F(5,130)=6.49$ ,  $p<.001$  in sixth graders) and version of the test booklet (Pillai's trace=0.10,  $F(5,163)=3.55$ ,  $p=.005$  in fourth graders; Pillai's trace=0.15,  $F(5,130)=4.56$ ,  $p<.001$  in sixth graders; note that there was no significant main effect found in fifth graders, Pillai's trace=0.03,  $F(5,151)=0.92$ ,  $p=.47$ ); that is, these two extraneous variables had partial effects on several indicators. There was no interaction between the school and booklet versions. Therefore, the school affiliation and version of the test booklet were included as control variables in the subsequent analysis.

### Partial correlations between the rate of diagram use in the four problems and each question type in the questionnaire

First, we conducted a partial correlation analysis to examine the degree of association between actual diagram use and each question format of the self-report questionnaires. Partial correlations between each format of the questionnaire and the average rate of diagram use in the test were calculated after controlling for school affiliations and versions of the test booklet for each grader (Tables 2, 3 and 4). All correlation coefficients were calculated by Pearson's method; the correlations between the average rate of diagram use in the test and whether each participant referred to diagram use in the free-description and multiple-choice questions were specially called the point-biserial correlation ( $r_{pb}$ ), which referred to a Pearson's correlation where one of the variables was dichotomous.

In the fourth grade, the rate of diagram use in the test was significantly positively correlated only with the mean frequency of subjective use of diagrams in the Likert scale ( $r=.34$ ,  $p<.001$ ) and not with the free description ( $r_{pb}=.07$ ,  $p=.38$ ) or multiple-choice ( $r_{pb}=.11$ ,  $p=.17$ ). In the fifth grade, there were significant positive correlations between the rate of diagram use in the test and the free description ( $r_{pb}=.22$ ,  $p=.006$ ), multiple-choice ( $r_{pb}=.25$ ,  $p=.001$ ), and Likert scale ( $r=.24$ ,  $p=.002$ ). In the sixth grade, as in the fifth grade, there were significant positive correlations between the rate of actual diagram use and each format of the self-reported measures (free description:  $r_{pb}=.34$ ,  $p<.001$ ; multiple-choice:  $r_{pb}=.23$ ,  $p=.008$ ; Likert scale:  $r=.37$ ,  $p<.001$ ).

In these correlation analyses, the four-point scale was used only for Likert-type questions, while the binary coding was used for free-description and multiple-choice questions to examine the association with actual strategy use. However, binarizing continuous responses are known to have lower statistical power than analyzing them continuously (e.g., Cohen, 1983; Fedorov et al., 2009; MacCallum et al., 2002). Therefore, the likelihood of significant correlations between Likert-type responses and the actual strategy use may have been affected by differences in statistical power with other question formats. Hence, as a supplementary analysis, we also confirmed the association between actual strategy use

**Table 2** Partial correlations between actual strategy use and each format of the self-report questionnaire in the fourth grade

	2	3	4
1. Rate of strategy use in the math test	.07	.11	.34***
2. Referring rate of diagram use in the free description	-	.45***	.29***
3. Choosing the rate of diagram use in the multiple-choice		-	.34***
4. Mean frequency of subjective diagram use in the Likert scale			-

Note. \*\*\*  $p < .001$

**Table 3** Partial correlations between actual strategy use and each format of the self-report questionnaire in the fifth grade

	2	3	4
1. Rate of strategy use in the math test	.22**	.25**	.24**
2. Referring rate of diagram use in the free description	-	.24**	.26***
3. Choosing the rate of diagram use in the multiple-choice		-	.32***
4. Mean frequency of subjective diagram use in the Likert scale			-

Note. \*\*  $p < .01$ , \*\*\*  $p < .001$

**Table 4** Partial correlations between actual strategy use and each format of the self-report questionnaire in the sixth grade

	2	3	4
1. Rate of strategy use in the math test	.34***	.23**	.37***
2. Referring rate of diagram use in the free description	-	.20*	.40***
3. Choosing the rate of diagram use in the multiple-choice		-	.40***
4. Mean frequency of subjective diagram use in the Likert scale			-

Note. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

and binarized Likert data. As a result, the correlation between actual strategy use and the binarized Likert answer had the same tendency as the four-point scale (see supplementary materials for details).

### Exploratory analysis: Regression analysis to examine the question format of the self-report questionnaire to predict actual strategy use

In addition to examining which question format of the self-report questionnaire was associated with actual diagram use in solving word problems, an exploratory multiple regression analysis was conducted to determine the unique contributions of each question format. The rate of actual diagram use in the test was an independent variable, and the reference rate of diagram use in the free description question, choice rate of the strategy of diagram use in the multiple-choice question, and subjective frequency of diagram use in the Likert-scale were dependent variables. Additionally, we included two variables as controls: school affiliation and the test booklet version. Among fourth graders, except for the version of the test booklet ( $b^* = -.23, p = .002$ ), only the Likert scale significantly predicted actual diagram use ( $b^* = .33, p < .001$ ; adjusted  $R^2 = .15, F(5, 165) = 6.87, p < .001$ ). However, fifth graders showed that multiple-choice questions significantly predicted actual strategy use ( $b^* = .17, p = .04$ );

the Likert scale ( $b^* = .15, p = .07$ ) and free description ( $b^* = .14, p = .09$ ) showed marginal significance for predicting actual diagram use (adjusted  $R^2 = .09, F(5, 153) = 3.96, p = .002$ ). In sixth graders, free description ( $b^* = .23, p = .01$ ) and Likert scale ( $b^* = .24, p = .009$ ) significantly predicted actual diagram use (adjusted  $R^2 = .21, F(5, 132) = 8.30, p < .001$ ).

This study examined the relationship between the rate of actual strategy use in four-word problems and the subjective indices of each question format, regardless of whether the answers were correct or not. However, diagram use is not required if word problems can be solved correctly without its use. Therefore, the responses of children who usually use diagrams appropriately but could solve the word problems without diagrams in this experimental test may have unfairly lowered the relationship between the actual and subjective diagram use. Hence, as a supplementary analysis, we also checked the relationship between the rate of diagram use and self-report questionnaires when limited to word problems answered incorrectly. A similar trend was still observed for all four-word problems in both correlations and regression analyses (see Tables S2–5 and the supplementary analysis sections in the supplementary materials).

## Discussion

This study examined the relationship between the actual strategic behavior of diagram use in solving mathematical word problems and participants' subjective reports through self-report questionnaires, including three types of question formats: free description, multiple-choice, and Likert-type questionnaires. We examined how each question format of the self-report measurement was associated with actual strategy use and the unique contribution each question format had in explaining actual strategy use as an exploratory analysis.

Our results indicated positive correlations between the self-report of subjective diagram use and actual behavior in the test, suggesting that self-report questionnaires can capture actual diagram use behavior to some extent, even in elementary school children. This aligns with Uesaka et al.'s (2007) findings of a moderate positive correlation between subjective self-report and actual behavior in junior high school students. Although previous studies have repeatedly noted that self-report questionnaires do not adequately reflect learners' actual strategy use (e.g., Desoete, 2008; Jacobse & Harskamp, 2012; Kikas & Jögi, 2016; McNamara, 2011; Saraç & Karakelle, 2012; Schellings & Van Hout-Wolters, 2011; Veenman, 2011), this study suggests that even with offline measures, appropriate answers can be obtained that reflect actual behavior depending on how the questions are asked, such as asking for free descriptions of their certain strategy use, presenting specific strategies and asking learners to choose one, and asking the frequency of use of each strategy on a scale.

Additionally, our exploratory results of regression analysis suggested that the appropriate question format depended on the grade level. Put differently, the variation of the question formats that better explain actual diagram use expands as grade increases; among the three formats, the Likert-type question showed a unique contribution to predicting the actual diagram use among all graders. In contrast, the free description questions on diagram use became more effective as the grade increased. Previous studies showed that the validity of self-report measures of metacognition was lower in elementary school children than in older students, suggesting a link with underdeveloped metacognitive skills (Craig et al., 2020). For younger children, especially those in the fourth grade, metacognitive skills are not yet

developed due to inexperience and insufficient expertise, and they cannot grasp their own strategy use properly. In contrast, older elementary school children, especially those in sixth grade, can monitor their strategy use and store it as knowledge and answer well even in the free-description format. This study suggested that while the Likert-type questionnaire is valid for all grades to some extent, methods with a higher degree of freedom, such as the free description type of the self-report questionnaire, become equally valid as students age. Our findings represent one possible way question formats can change the way of capturing actual strategy use well. Further examination is needed in future studies.

The result that the predictability of actual strategy use differed depending on the question format in the self-report questionnaire may reflect differences in the difficulty of remembering one's own experience depending on the form of recall or recognition. When remembering something, it has been shown that performance on the recognition task is generally better than that of the recall task (Postman et al., 1975). In other words, when comparing the multiple-choice and Likert questions, presented as option items describing specific strategies as in the recognition task, and the free description question, presented to remember and freely write down individuals' experiences regarding strategy use, such as the recall task, it is possible that recognizing one's experience is easier even for younger children because the options in the questionnaire may work as a remembrance cue. In addition, although both the multiple-choice and Likert scales present items indicating specific strategies, the Likert type, which asks participants to rate the frequency of each strategy use separately (i.e., similar to the yes/no recollection task), can provide a more detailed step-by-step remembering of the children's experience of strategy use than the multiple-choice type, which asks them to select an appropriate item from a list of candidate items (like the batch test). This may be why the Likert response in fourth graders was only able to predict strategy use behavior.

Previous studies have repeatedly suggested that offline ratings using a Likert scale to ask for the frequency of strategy use do not adequately reflect actual strategy use behavior (e.g., Jacobse & Harskamp, 2012). Nevertheless, there was a certain degree of correlation between the Likert scale and actual behavior in this study. This may have been influenced by our target strategy for asking about the frequency of use, that is, the strategy of diagram use. This strategy involved an overt action of whether to write a diagram or not. Therefore, it was easy for elementary school children to judge whether they usually used the diagram strategy, and there could be a positive association between actual behavior and self-reporting.

Furthermore, our results also indicated the possibility that a combination of several question formats provides a more effective and valid measurement of strategy use. This study's exploratory analysis showed the unique contribution of each question format to the prediction of actual strategy use. Most previous studies measuring learners' strategy use using self-report questionnaires tended to evaluate it only through the Likert format (e.g., Bråten & Olaussen, 1998; Cano, 2006; Duncan & McKeachie, 2005; Magno, 2011; Pintrich et al., 1993; Sungur, 2007; Weinstein et al., 1988); however, it is possible that a combination of free description measures, in addition to the Likert scale, provided a more valid measure, at least for sixth graders and above. However, since our results were obtained from an exploratory analysis of a limited sample, further investigation is needed.

## Study limitations and future directions

While we showed that actual strategy use behavior could be predicted by self-report offline measures in elementary school children through adaptation of the question format to the grade level, several limitations remain. First, this study targeted only one learning strategy. In this study, the relationship between actual strategy use and self-report was examined only for the diagram use strategy. Learning strategies do not always involve externally observable behaviors. If we examine the relationship between online and offline measures for learning strategies such as elaboration, which does not involve explicit behavior, we can confirm whether the self-report questionnaire with an appropriately chosen format can truly predict actual strategy use even in fourth graders, as in the present study.

The second limitation was the instability of the multiple-choice question. It has been suggested that when participants are asked to choose among several options, the target's choice behavior is influenced by the repertoire of options other than the target (known as a decoy effect, e.g., Wedell & Pettibone, 1996). In this study, the contents of the options were decided through a consensus of the authors; however, we cannot deny a bias possibility that made certain options more or less likely to be selected. Future studies must examine whether the choices were appropriate and pay special attention to the bias in the ease of selection by choice options.

Third, the age range in this survey was another limitation. We surveyed children who had just entered grades 4–6. To examine the developmental effects in more detail, a cross-sectional survey including younger children must be conducted.

Finally, the situational dependence of strategy use should be considered in future studies. In this study's questionnaire, many children responded to more than one strategy in their free-description answer. It is natural for strategy use to vary depending on the situation, context, and problem content at the time. Hence, it remains unclear what particular problem or context the children imagined when they answered the questionnaire. Future research should consider the extent of strategy use by being more specific about context and problem content and asking about the order in which particular strategies were recalled from the strategy repertoire.

In summary, to clarify whether or how to properly measure actual strategy use with offline measurements, this study examined the impact of the question format of self-report questionnaires on whether it better predicted actual strategy use. Hence, we targeted elementary school children, for whom the relationship between offline-online indicators is considered particularly low. The results showed that some question types could predict strategy use behavior to a certain degree and that the appropriate question types varied depending on the grade level. Future studies must set up an appropriate question format according to the age of the respondents when measuring strategy use with self-report questionnaires.

## Additional information

### Current themes of research

Tatsushi Fukaya: (1) Developing effective interventions to enhance students' metacognition and use of learning strategies, (2) Developing teacher training methods to promote the

acquisition of their professional knowledge. Shiho Kashihara: Memory distortion, agency, metacognition, and the neural basis of such human cognition.

## Relevant publications

Fukaya, T., & Uesaka, Y. (2019). Cultivating pre-service and in-service teachers' abilities to deepen understanding and promote learning strategy use in pupils. In E. Manalo (Ed.), *Deeper learning, dialogic learning, and critical thinking* (pp. 279-296). Routledge.

Fukaya, T., Uesaka, Y., & Ichikawa, S. (2019). Investigating the effects of Thinking after Instruction approach: An experimental study of science class. *Educational Technology Research, 41*(1), 1-11. <https://doi.org/10.15077/etr.42105>

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10212-022-00668-z>.

**Acknowledgements** We are grateful to Mami Nishidai for her help with data coding.

**Authors' contributions** All authors contributed to the study's conception and design, material preparation, data collection, and analysis. The first draft of the manuscript was written by Shiho Kashihara and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** This work was supported by JSPS KAKENHI Grant Number 19H01756.

**Data availability** Not applicable.

**Code availability** Not applicable.

## Declarations

**Competing interests** The authors declare that they have no conflict of interest.

**Ethics approval** The study was approved by the Ethical Committee of Hiroshima University.

**Consent** We obtained informed consent from the school principals for the children to participate in this study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bayard, N. S., van Loon, M. H., Steiner, M., & Roebers, C. M. (2021). Developmental improvements and persisting difficulties in children's metacognitive monitoring and control skills: cross-sectional and longitudinal perspectives. *Child Development, 92*(3), 1118–1136. <https://doi.org/10.1111/cdev.13486>.
- Berger, J. L., & Karabenick, S. A. (2016). Construct validity of self-reported metacognitive learning strategies. *Educational Assessment, 21*(1), 19–33. <https://doi.org/10.1080/10627197.2015.1127751>.



- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417–444. <https://doi.org/10.1146/annurev-psych-113011-143823>.
- Bråten, I., I., & Olaussen, B. S. (1998). The relationship between motivational beliefs and learning strategy use among norwegian college students. *Contemporary Educational Psychology*, 23(2), 182–194. <https://doi.org/10.1006/ceps.1997.0963>.
- Cano, F. (2006). An in-depth analysis of the learning and study strategies inventory (LASSI). *Educational and Psychological Measurement*, 66(6), 1023–1038. <https://doi.org/10.1177/0013164406288167>.
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7(3), 249–253. <https://doi.org/10.1177/014662168300700301>.
- Craig, K., Hale, D., Grainger, C., & Stewart, M. E. (2020). Evaluating metacognitive self-reports: systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning*, 15(2), 155–213. <https://doi.org/10.1007/s11409-020-09222-y>.
- Cromley, J., & Azevedo, R. (2011). Measuring strategy use in context with multiple-choice items. *Metacognition and Learning*, 6(2), 155–177. <https://doi.org/10.1007/s11409-011-9070-z>.
- Dermitzaki, I. (2005). Preliminary investigation of relations between young students' self-regulatory strategies and their metacognitive experiences. *Psychological Reports*, 97(3), 759–768. <https://doi.org/10.2466/pr0.97.3.759-768>.
- Derry, S. J., & Murphy, D. A. (1986). Designing systems that train learning ability: from theory to practice. *Review of Educational Research*, 56(1), 1–39. <https://doi.org/10.3102/00346543056001001>.
- Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: how you test is what you get. *Metacognition and Learning*, 3(3), 189–206. <https://doi.org/10.1007/s11409-008-9026-0>.
- Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology*, 47(2), 274–296. [https://doi.org/10.1016/0022-0965\(89\)90033-7](https://doi.org/10.1016/0022-0965(89)90033-7).
- Duncan, T. G., & McKeachie, W. J. (2005). The making of the motivated strategies for learning Questionnaire. *Educational Psychologist*, 40(2), 117–128. [https://doi.org/10.1207/s15326985sep4002\\_6](https://doi.org/10.1207/s15326985sep4002_6).
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>.
- Edwards, A. J., Weinstein, C. E., Goetz, E. T., & Alexander, P. A. (2014). *Learning and study strategies: issues in assessment, instruction, and evaluation*. Elsevier.
- Fedorov, V., Mannino, F., & Zhang, R. (2009). Consequences of dichotomization. *Pharmaceutical Statistics*, 8(1), 50–61. <https://doi.org/10.1002/pst.331>.
- Haberkorn, K., Lockl, K., Pohl, S., Ebert, S., & Weinert, S. (2014). Metacognitive knowledge in children at early elementary school. *Metacognition and Learning*, 9(3), 239–263. <https://doi.org/10.1007/s11409-014-9115-1>.
- Hembree, R. (1992). Experiments and relational studies in problem solving: a meta-analysis. *Journal for Research in Mathematics Education*, 23(3), 242–273. <https://doi.org/10.5951/jresmetheduc.23.3.0242>.
- Jacobse, A. E., & Harskamp, E. G. (2012). Towards efficient measurement of metacognition in mathematical problem solving. *Metacognition and Learning*, 7(2), 133–149. <https://doi.org/10.1007/s11409-012-9088-x>.
- Jönsson, A., Rosenlund, D., & Alvé, F. (2017). Complement or contamination: A study of the validity of multiple-choice items when assessing reasoning skills in physics. *Frontiers in Education*, 2. <https://doi.org/10.3389/educ.2017.00048>
- Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: do students practise retrieval when they study on their own? *Memory (Hove, England)*, 17(4), 471–479. <https://doi.org/10.1080/09658210802647009>.
- Kikas, E., & Jögi, A. L. (2016). Assessment of learning strategies: self-report questionnaire or learning task. *European Journal of Psychology of Education*, 31(4), 579–593. <https://doi.org/10.1007/s10212-015-0276-3>.
- Lehmann, M., Hagen, J., & Ettinger, U. (2022). Unity and diversity of metacognition. *Journal of Experimental Psychology: General*, 151(10), 2396–2417. <https://doi.org/10.1037/xge0001197>.
- Lowrie, T. (2020). The utility of diagrams in elementary problem solving. *Cognitive Development*, 55, 100921. <https://doi.org/10.1016/j.cogdev.2020.100921>.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19–40. <https://doi.org/10.1037/1082-989x.7.1.19>.
- Magno, C. (2011). Validating the academic self-regulated learning scale with the motivated strategies for learning questionnaire (MSLQ) and learning and study strategies inventory (LASSI). *The International Journal of Educational and Psychological Assessment*, 7(2), 56–73. <https://ssrn.com/abstract=2287180>.

- McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: Challenges and successes. *Metacognition and Learning*, 6(2), 195–203. <https://doi.org/10.1007/s11409-011-9082-8>.
- Metcalfe, J., & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning*, 8(1), 19–46. <https://doi.org/10.1007/s11409-013-9094-7>.
- Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40. <https://doi.org/10.1037//0022-0663.82.1.33>.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, 53(3), 801–813. <https://doi.org/10.1177/0013164493053003024>.
- Postman, L., Kruesi, E., & Regan, J. (1975). Recognition and recall as measures of long-term retention. *The Quarterly Journal of Experimental Psychology*, 27(3), 411–418. <https://doi.org/10.1080/14640747508400501>.
- R Core Team. (2021). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, J. T. E. (2004). Methodological issues in questionnaire-based research on student learning in higher education. *Educational Psychology Review*, 16(4), 347–358. <https://doi.org/10.1007/s10648-004-0004-z>.
- Rodriguez, M. C. (2003). Construct equivalence of multiple-choice and constructed-response items: a random effects synthesis of correlations. *Journal of Educational Measurement*, 40(2), 163–184. <https://doi.org/10.1111/j.1745-3984.2003.tb01102.x>.
- Roebbers, C. M., Krebs, S. S., & Roderer, T. (2014). Metacognitive monitoring and control in elementary school children: their interrelations and their role for test performance. *Learning and Individual Differences*, 29, 141–149. <https://doi.org/10.1016/j.lindif.2012.12.003>.
- Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *The British Journal of Educational Psychology*, 77(2), 351–378. <https://doi.org/10.1348/000709906X106147>.
- Saraç, S., & Karakelle, S. (2012). On-line and off-line assessment of metacognition. *International Electronic Journal of Elementary Education*, 4(2), 301–315. <https://www.iejee.com/index.php/IEJEE/article/view/201>.
- Schellings, G. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning*, 6(2), 91–109. <https://doi.org/10.1007/s11409-011-9069-5>.
- Schellings, G., & Van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: theoretical and empirical considerations. *Metacognition and Learning*, 6(2), 83–90. <https://doi.org/10.1007/s11409-011-9081-9>.
- Schellings, G. L. M., van Hout-Wolters, B. H. A. M., Veenman, M. V. J., & Meijer, J. (2013). Assessing metacognitive activities: the in-depth comparison of a task-specific questionnaire with think-aloud protocols. *European Journal of Psychology of Education*, 28, 963–990. <https://doi.org/10.1007/s10212-012-0149-y>.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.
- Sungur, S. (2007). Modeling the relationships among students' motivational beliefs, metacognitive strategy use, and effort regulation. *Scandinavian Journal of Educational Research*, 51(3), 315–326. <https://doi.org/10.1080/00313830701356166>.
- Tullis, J. G., & Maddox, G. B. (2020). Self-reported use of retrieval practice varies across age and domain. *Metacognition and Learning*, 15(2), 129–154. <https://doi.org/10.1007/s11409-020-09223-x>.
- Uesaka, Y., Manalo, E., & Ichikawa, S. (2007). What kinds of perceptions and daily learning behaviors promote students' use of diagrams in mathematics problem solving? *Learning and Instruction*, 17(3), 322–335. <https://doi.org/10.1016/j.learninstruc.2007.02.006>.
- Veenman, M. V. J. (2011). Alternative assessment of strategy use with self-report instruments: a discussion. *Metacognition and Learning*, 6(2), 205–211. <https://doi.org/10.1007/s11409-011-9080-x>.
- Veenman, M. V. J., & van Cleef, D. (2019). Measuring metacognitive skills for mathematics: students' self-reports versus on-line assessment methods. *ZDM: The International Journal on Mathematics Education*, 51(4), 691–701. <https://doi.org/10.1007/s11858-018-1006-5>.
- Wedell, D. H., & Pettibone, J. C. (1996). Using judgments to understand decoy effects in choice. *Organizational Behavior and Human Decision Processes*, 67(3), 326–344. <https://doi.org/10.1006/obhd.1996.0083>.
- Weinstein, C. E., Zimmermann, S. A., & Palmer, D. R. (1988). Assessing learning strategies: The design and development of the LASSI. In C. E. Weinstein, E. T. Goetz, & P. A. Alexander (Eds.), *Learning and Study Strategies* (pp. 25–40). Academic Press. <https://doi.org/10.1016/B978-0-12-742460-6.50009-8>

---

Zimmerman, B. J., & Martinez-Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology*, *80*(3), 284–290. <https://doi.org/10.1037/0022-0663.80.3.284>.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.