JOURNAL
OF BALTIC
SCIENCE
EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

DEVELOPING AND VALIDATING AN INSTRUMENT TO ASSESS NINTH-GRADE STUDENTS' ONLINE METACOGNITIVE SKILLS IN SOLVING CHEMISTRY PROBLEMS

Yating Zeng,
Shaohui Chi,
Zuhao Wang,
Xiaosong Zhuang

Abstract. *Online metacognitive skills are the real-time awareness of cognition, which can effectively promote science learning and improve performance in solving scientific problems. Therefore, it is important to enhance and diagnose students' online metacognitive skills in science education. This study aimed to evaluate ninth-grade students' online metacognitive skills while processing chemistry problems. To achieve this goal, this study constructed a framework for guiding the development of an instrument comprising 12 two-tier items. A total of 258 ninth graders took part in the field testing in Jiangsu, China. A partial credit Rasch model analysis was employed to inform instrument development and evaluation. The results revealed that this instrument was valid and reliable for assessing students' online metacognitive skills. Nearly 60% of the ninth-grade students in this sample were able to monitor their own thought processes or evaluate their own cognitive performance in processing chemistry problems. About one-third of the students could regulate their thought processes. However, less than 4% of the students could make attributions about their cognitive performance.*

Keywords: *assessment instrument, problem-solving skills, Rasch measurement model, chemistry education*

Yating Zeng, Shaohui Chi,
Zuhao Wang, Xiaosong Zhuang
East China Normal University, China

Introduction

Metacognition refers to the knowledge and awareness of cognition (Flavell, 1979), indicating a person's ability to reflect upon their own thoughts, experiences, and actions (Soto et al., 2018; Weil et al., 2013). Previous literature has proposed two types of metacognition: offline metacognition and online metacognition (De Clercq et al., 2000; Desoete, 2001). Offline metacognition is a general reflection on past or future activities (Violeau et al., 2020). In contrast, online metacognition is a mental process activated as the individual is processing the task at hand (Ng et al., 2021; Quiles et al., 2014). The term *online*, here, is a synonym for an ongoing situation where certain metacognitive functions take place (Mazancieux et al., 2021).

In science education, online metacognition has been considered more important than offline metacognition, specifically in solving scientific problems (Gilbert, 2005). The use of online metacognition can facilitate students' abilities to actively control and regulate their thinking process for processing tasks (Efklides & Misailidi, 2010; Goldstein & Naglieri, 2011; Kuzle, 2018; Lavi et al., 2019; Seel, 2012; She et al., 2012). For instance, Cooper et al. (2008) found that the learners who performed well on chemistry problems scored high on measures of online metacognition. Rickey and Stacy (2000) also stated that online metacognition could compensate for lacking chemical problem-solving experience, and the students who were aware of and in control of their own thoughts were better able to solve chemistry problems than the students who were not online metacognitive.

A large number of studies, however, have shown that students have limited online metacognitive skills in solving chemistry problems, as measured primarily by observation, verbal protocols or self-reports (Mathabathe & Potgieter, 2017; Sandi-Urena et al., 2011; Wang, 2022). For instance, Pulmones (2007) discovered that many students were unaware of their thoughts, as

https://doi.org/10.33225/jbse/23.22.520

seen by their failure to monitor the progress of their work, determine whether their solutions were accurate, and explain why the problems were challenging. According to Bell and Volckmann (2011), at least half of the students overestimated their performance, and the worse they performed, the more biased they were. Teichert et al. (2017) noticed that students struggled with online metacognitive reflection on their problem-solving processes, and they had trouble explaining their problem-solving results.

Therefore, there is an urgent need to develop students' online metacognitive skills to help them to become successful problem solvers and obtain their expected learning outcomes. To achieve this goal, one essential step towards fostering students' online metacognitive skills is to diagnose the extent of students' online metacognitive skills in solving chemistry problems. However, to date, limited studies have focused on assessing students' specific online metacognitive skills in solving chemistry problems (Schellings, 2011; Wang, 2015). In this regard, the present study aimed to evaluate ninth-grade students' online metacognitive skills in solving chemistry problems. To fulfill this purpose, this study constructed a framework to guide the development of an instrument for identifying the specific online metacognitive skills that are challenging for students, which in turn would allow appropriate teaching aids to be provided.

Literature Review

Online Metacognitive Skills

In science education, current literature has not achieved a consensus regarding the construct of online metacognitive skills (Desoete, 2001; Desoete & Roeyers, 2002; Efklides, 2002; Lee et al., 2019). According to Chen and Goverover (2021), online metacognition skills comprise the abilities of monitoring (e.g., identifying mistakes) and regulation (e.g., correcting responses). Similarly, Quiles et al. (2014; 2020) have identified two aspects of online metacognitive skills, including online metacognitive monitoring that involves the subjective self-assessment of confidence in answering the questions and online metacognitive control concerning the decision on whether to validate the answers. Sporer and Horry (2011) have noted that online metacognitive evaluation is an essential ability to make estimates of the likely accuracy of decisions. In addition, researchers have also proposed that online metacognitive skills consist of students' abilities to make attributions for the success or failure of their performance when confronted with problems (Wong, 2007) and their consciousness of task processing (Efklides et al., 1998; 2008). In this sense, online metacognitive skills is an umbrella term encompassing the interrelated sub-skills necessary for solving problems.

Online Metacognitive Skills in Solving Chemistry Problems

Online metacognitive skills are critical in tackling chemistry problems (Wang, 2015). Students who employed more efficient online metacognitive strategies, such as using regulatory skills and reflective practices, could handle more complex chemistry problems (Cooper et al., 2008; Sandi-Urena et al., 2011). A study conducted by She et al. (2012) has shown that the top performers in solving chemistry problems were good at regulating their actions and evaluating whether the answers were correct via checking information. Teichert et al. (2017) found that accurate monitoring and reflection on ongoing mental processes could facilitate students' application of cognitive resources within chemistry problem contexts.

Specifically, Heidbrink and Weinrich (2021) proposed three iterative phases regarding the online metacognitive process in solving a chemistry problem. While solving a chemistry problem, students first identify the given information. They then deal with the information provided to ensure that it is working towards the goal, which involves being aware of some mistakes and debugging the errors. Finally, students reflect on the whole process. Similarly, Kipnis and Hofstein's (2008) study has demonstrated that students judge whether the solutions to problems are logical by carefully checking their actions and asking themselves if they are doing the right thing and if the results are reasonable.

Assessing Online Metacognitive Skills

In the field of chemistry education, students' metacognitive skills have been assessed in a variety of ways. For instance, Cooper and Sandi-Urena (2009) developed a scale where students rated their degree of agreement with

https://doi.org/10.33225/jbse/23.22.520

27 metacognitive statements (e.g., I do not check if the answers are reasonable). Hawker et al. (2016) and Testa et al. (2023) asked students to estimate their performance after examinations and then computed the discrepancies between the actual and predicted scores. Gamby and Bauer (2022) evaluated metacognitive awareness and skills by examining students' comments on their cognitive knowledge and beliefs.

However, many researchers have warned that the previous measurements of metacognitive skills cannot assess online metacognitive skills because those measurements are unreliable, as they often fail to capture the dynamic metacognitive processes within a particular task (Jacobse & Harskamp, 2012; Wang, 2015). The validity of the online metacognitive data not gathered by online methods should also be questioned (Schellings, 2011).

To evaluate students' online metacognitive skills, some researchers have proposed so-called online measurement techniques (Bryce & Whitebread, 2012; Veenman et al., 2006), including think-aloud, observation, and eye-tracking methods administered to students during actual task performance (Bannert & Mengelkamp, 2008; Kinnunen & Vauras, 2010; Veenman, 2013; Winne, 2014). For instance, Dermitzaki (2005) conducted a direct observation to investigate second graders' self-regulative behaviour during the construction of a wooden vehicle or toy.

Although online measurement techniques have shed light on detecting students' online metacognitive skills, the majority of those techniques are labour- and time-intensive (Bannert & Mengelkamp, 2008; Veenman & van Cleef, 2019). For example, the think-aloud method requires well-trained raters to put a huge effort into scoring verbal protocols (Schellings et al., 2013). Observation of online metacognitive behaviour is time-consuming (Veenman & van Cleef, 2019) and often requires a substantial amount of effort to collect data, and the behavioural data must be examined by several observers according to a complex scoring system (McCord & Matusovich, 2019; Veenman, 2017).

Compared with other online measurement techniques, questionnaires are the least labour-consuming method of acquiring and analyzing a massive amount of information (Schellings et al., 2013), and have therefore been widely used to measure online metacognitive skills. For instance, Lawanto (2010) developed the Engineering Design Project Inventory to measure university students' online metacognitive skills in the process of engineering design. All data from the questionnaire were collected from participants during the engagement process.

Specifically, two-tier items, consisting of one task and one question eliciting the online metacognitive activities behind the task response, have been one of the most popular types of questionnaire items for efficiently investigating online metacognitive skills (Dermitzaki, 2005). For example, Koren et al. (2005) and Quiles et al. (2014) used two-tier items (the Wisconsin Card Sorting Test, WCST) to evaluate online metacognitive monitoring and control. The participants first performed a matching task. Then, they were asked to rate their confidence in their responses, from which the extent of online metacognitive monitoring was measured. The participants were also requested to validate the task responses and decide whether the responses should count towards their overall grade, thus examining their online metacognitive control.

In short, previous studies have made valuable contributions to investigating students' online metacognitive skills. However, the currently articulated online metacognitive skills in solving chemistry problems have received too little attention in the research literature. While primary, upper-secondary, and university students were the main subjects in most of the existing related studies, the extent of lower-secondary school students' online metacognitive skills in dealing with chemistry problems remains unknown. In order to address this gap, the present study aimed to develop and validate a measurement instrument by applying Wilson's Construct Modeling Approach (Wilson, 2005). The findings of this study may provide a basis for further instruction regarding online metacognition in solving chemistry problems.

Research Purpose and Research Questions

The current study intended to develop and validate a measurement instrument to assess ninth-grade students' online metacognitive skills in solving chemistry problems.

Specifically, two research questions were addressed:

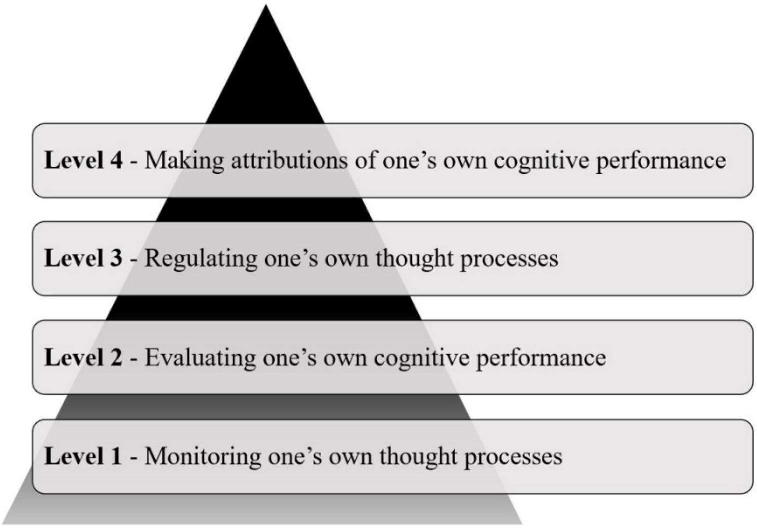
1. What evidence supports the reliability and validity of measures of the instrument developed in this study for assessing students' online metacognitive skills in solving chemistry problems?
2. What is the extent of ninth-grade students' ability to monitor, evaluate, regulate, and make attributions of their own cognitive performance?

https://doi.org/10.33225/jbse/23.22.520

Theoretical Framework

In reference to previous studies (Chen & Goverover, 2021; Quiles et al., 2020; Sporer & Horry, 2011; Wong, 2007), this study defined online metacognitive skills as the ability to monitor, evaluate, regulate, and reflect on one's own performance in solving chemistry problems. In particular, four hierarchical levels comprise the construct of this study, which is hypothesized to delineate a set of complicated online metacognitive skills, from the least sophisticated (i.e., monitoring one's own thought processes) to the most advanced (i.e., making attributions about one's own cognitive performance), as shown in Figure 1.

Figure 1
Online Metacognitive Skills Framework



The first level relates to students' ability to monitor their thoughts. Monitoring has long been emphasized in scientific problem-solving literature (Hollingworth & McLoughlin, 2001; Rozencwajg, 2003). Monitoring processes frequently occur in the early stages of problem-solving and play essential roles in helping students modify their behaviour (Jacobse & Harskamp, 2012; Kuzle, 2013; Rickey & Stacy, 2000). Online metacognitive monitoring processes require students to concentrate on ongoing task performance, recognize if the thinking is clear, and identify whether the reasoning is correct or incorrect to ensure that the current thinking is going in the right direction.

The second level concentrates on students' ability to evaluate their cognitive performance, asking students to demonstrate the ability to assess their problem-solving processes, estimate the correctness and rationality of the solutions, and check their answers. Evaluation has been widely considered an essential element of online metacognitive skills that helps verify reasonable solutions to chemistry problems (Vo et al., 2022). Given that making performance judgments often occurs after and is dependent on students' monitoring behaviour (Jacobse & Harskamp, 2012; Koriat, 2002), evaluating one's cognitive performance is more demanding for students compared with monitoring (Azizah et al., 2019).

The third level relates to students' abilities to regulate their thought processes. Students who achieve this level are able to find and correct their mistakes, modify their operations, overcome their difficulties, and develop new solutions to deal with problems in an efficient way. As a critical component of online metacognitive skills (Rickey & Stacy, 2000), regulation is more advanced than monitoring or evaluating because it requires students to accurately monitor and assess their performance before they can direct their cognition (Cheng, 2011).

The fourth level corresponds to students' ability to make attributions of their own cognitive performance, requiring students to learn from their successes and failures by reflecting on their problem-solving processes and identifying the factors affecting their performance. Making attributions is not only important for solving chemistry problems (Zoller & Pushkin, 2007), but it can also influence judgment, self-regulation process, and achievement striving (Clifford, 1986; Lin, 2003; Ross, 1999). The ability to make attributions for their performance depends on abilities such as monitoring, evaluation, and regulation (Garner, 2009; Zimmerman, 2011; 2013). Therefore, this ability is considered the most advanced in terms of online metacognitive skills.

Research Methodology

General Background

Until now, few studies have focused on evaluating lower-secondary school students' online metacognitive skills in processing chemistry problems. Therefore, the present study developed and validated an instrument to assess ninth-grade students' online metacognitive skills in solving chemistry problems. The validity and reliability of the instrument were examined using Rasch analysis. Students' online metacognitive skills were detected during the 2020–2021 academic year.

Sample Selection

Convenience sampling was used to recruit participants from two lower-secondary schools in Jiangsu, China. For the pilot test, 236 ninth graders (130 females and 106 males) were invited from one school to evaluate and refine the initial instrument. For the field test, 258 ninth graders (123 females and 135 males) were selected from the other school to confirm the revised instrument. The sample size was determined based on previous studies, which suggested that at least 200 participants would be adequate (Liu, 2020). These students were academically diverse and included low-, medium-, and high-performers. All the participants were volunteers and were assured of confidentiality.

Instrument and Procedures

The instrument development was guided by Wilson's Construct Modeling Approach (2005) through six iterative steps: (1) developing a construct map; (2) generating items; (3) developing scoring rubrics; (4) administering a pilot test and revising items; (5) conducting the field test; and (6) Rasch analysis.

Developing a construct map. A construct map specifies a continuum of skills that learners are expected to progress through for the target construct (Rittle-Johnson et al., 2011). In this study, a construct map (see Figure 1) for online metacognitive skills in chemistry problem-solving was developed.

Generating items. In line with previous studies (e.g., Dermitzaki, 2005), a two-tier response format was adopted to generate the test items. The initial item pool was generated based on a comprehensive review of related measurements, including the Engineering Design Project Inventory (Lawanto, 2010) and the Wisconsin Card Sorting Test (Koren et al., 2005). A group of experts (three science education professors and two experienced science educators) and the authors took part in the development of the items.

In particular, the following three tenets were used to review and revise the items in the initial pool: (1) each item should be linked to one level of skills, and all levels should have corresponding items. (2) The chemistry problems should be embedded in real-life contexts and could not be taken from the textbook, thereby avoiding the possibility of students solving problems by memorizing, and the problems must be cognitively challenging to promote students to think deeply and stimulate their online metacognitive thinking processes. (3) Given that the test aimed to assess online metacognitive skills, the knowledge of chemistry required should be reduced to a minimum.

The initial version of the instrument included 20 two-tier items embedded in five tasks. An example task and details of the item illustrations are presented in Appendix A.

Developing scoring rubrics. After the items were designed, scoring rubrics were constructed to evaluate students' responses to the items. In this study, the scoring rubrics contained information about the performance indicators and codes for identifying the anticipated responses corresponding to the related level of performance. Among the initial 20 items, five items were scored from 0 to 1, and the other 15 items were scored from 0 to 2 (examples are shown in Appendix B).

Four graduate students in science education were recruited as raters to score each item independently. All the discrepancies that arose during the rating processes were resolved through several extensive discussions. The inter-rater reliability was determined as Cohen's kappa value of .95 ($p < .01$).

https://doi.org/10.33225/jbse/23.22.520

Administering a pilot test and revising pilot test items. To further refine the instrument, participants not only took part in the pilot test but also were interviewed afterwards to collect information regarding their experiences in the test. The results of the pilot test were analyzed using Rasch analysis to check the quality of the items. Based on the pilot test results, two tasks were removed because of low item discrimination, and three controversial items were modified following a focus group discussion with experts and student interviews.

The final version of the instrument comprised 12 two-tier items. Table 1 shows the distribution of the items corresponding to the different levels of online metacognitive skills in solving chemistry problems.

Table 1
Field Item Distribution Regarding Online Metacognitive Skills in Solving Chemistry Problems

Levels	Performance expectations	Items
Level 1	Monitoring one's own thought processes	Q2/Q3; Q8/Q9; Q14/Q15
Level 2	Evaluating one's own cognitive performance	Q2/Q5; Q8/Q11; Q14/Q17
Level 3	Regulating one's own thought processes	Q1/Q4; Q7/Q10; Q13/Q16
Level 4	Making attributions of one's own cognitive performance	Q5/Q6; Q11/Q12; Q17/Q18

Conducting the field test. A 40-minute field test was conducted to ensure the validity and reliability of the revised instrument, as well as to determine the online metacognitive skills within the sample. The Rasch model was employed to analyze the field test data.

Rasch analysis. In science education, Rasch measurement has been widely used for the rigorous development and examination of measurement instruments (Boone & Scantlebury, 2006; Planinic et al., 2019; Sideridis, 2007; Wang et al., 2017; Wren & Barbera, 2014). This study conducted Rasch analysis taking into account the following benefits: (a) the Rasch model is expressed both at the instrument level and the item level (Muller, 2005). (b) Rasch measurement focuses on the likelihood of the response rather than the answer per se (Wilson et al., 2006). (c) The features of the participants do not affect the item estimates (Chae et al., 2018). (d) Rasch measurement allows for converting the raw ordinal scores into interval equivalent scores expressed on a linear scale (Engelhard & Myford, 2003).

Data Analysis

The development of the assessment instrument for this study and subsequent analysis of pilot and field test results were all guided by the application of Rasch measurement. Given that all the items in this study did not share the same scale steps, partial credit Rasch model analysis was applied for measurement development and instrument evaluation. All raw data were converted into Rasch measures (in logits) to generate interval data for further statistical calculations (Bond & Fox, 2015). Specifically, the person/item separation and person/item reliability indices were calculated to examine the instrument's reliability. Fit statistics and the person-item distribution map (Wright map) were used to evaluate the validity of the instrument.

Research Results

In presenting the findings, this study focused on the revised version of the measurement instrument. Data from the field testing were analyzed using Winsteps software (Linacre, 2009) version 3.72.3.

Evidence for Reliability and Validity

Unidimensionality. Unidimensionality is an explicit requirement for Rasch measurement (Smith, 1996), assuming that the test measures only one underlying construct (Wind & Schumacker, 2021). This study performed a principal component analysis of residuals to check for unidimensionality (Bond & Fox, 2015). According to Linacre (2020), unidimensionality is achieved if the unexplained variance in the first contrast is less than two eigenvalues. The results indicated that the eigenvalue was 1.9 in the first contrast, thus supporting unidimensionality.

https://doi.org/10.33225/jbse/23.22.520

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Reliability. In Rasch measurement, item reliability refers to the consistency of relative item measure location, and person reliability refers to the replicability of person placement across items (Bond & Fox, 2015; Linacre, 2020). The Rasch measurement model provides indices of item/person separation and item/person reliability like Cronbach's alpha for evaluating reliability across persons and items. The criteria for accepting reliability and separation in the Rasch model are above .70 (Fitzpatrick et al., 2004) and 2, respectively (Fisher, 2007). As Table 2 shows, the item reliability was high at .98, and the person reliability was .71. However, the item separation was acceptable at 7.96, and the person separation was low at 1.55.

Table 2

Field Test Summary Statistics from Rasch Model Analysis

	Measure	Error	Infit		Outfit		Separation	Reliability
			MNSQ	ZSTD	MNSQ	ZSTD		
Person	−0.24	0.61	1.00	−0.10	1.02	−0.10	1.55	.71
Item	0.00	0.11	0.99	−0.50	1.02	0.20	7.96	.98

Validity. Rasch fit statistics and the person-item distribution map (Wright map) are used to test the validity of the estimated measures (Bond & Fox, 2015). The Rasch measurement model generates a set of fit statistics to evaluate the extent to which each item matches the expectations of the Rasch model. This study reviewed Outfit/Infit mean square residual (MNSQ) values and standardized mean square residual (ZSTD) values, with MNSQ values between 0.5 and 1.5 and ZSTD values between −2 and +2 as fit criteria for quality assurance (Bond & Fox, 2015). The point-measure correlation index (PTMEA) was also checked. The PTMEA refers to the Pearson correlation between the item score and the Rasch measure (Linacre, 2020), ranging from −1.0 to 1.0. The expected point-measure correlation value is positive, showing that the item-level scoring accords with the latent variable (Bond & Fox, 2015). As Table 3 shows, all the field items had acceptable MNSQ (ranging from 0.57 to 1.43) and PTMEA values (ranging from .39 to .63). However, nine of the twelve two-tier items had ZSTD values outside the fit range.

Table 3

Fit Statistics of 12 Two-tier Field Items from Rasch Model Analysis

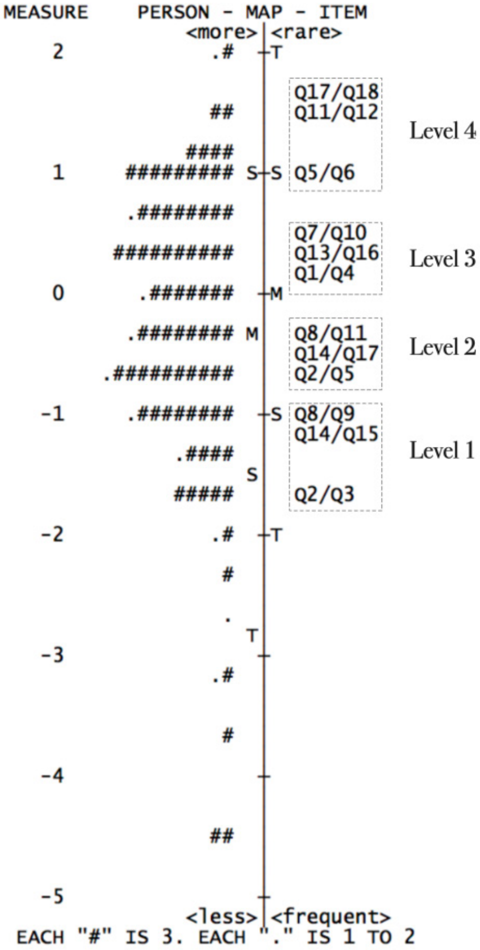
Item	Measure	Model S.E.	Infit		Outfit		PT-MEASURE CORR.
			MNSQ	ZSTD	MNSQ	ZSTD	
Q17/Q18	1.60	.13	0.63	−5.6	0.65	−3.2	.60
Q11/Q12	1.46	.13	0.57	−6.5	0.59	−4.1	.63
Q5/Q6	1.03	.12	0.60	−5.7	0.76	−2.4	.55
Q7/Q10	0.51	.12	1.26	2.8	1.43	3.8	.40
Q13/Q16	0.38	.12	1.20	2.2	1.37	3.4	.39
Q1/Q4	0.23	.12	1.17	1.8	1.23	2.2	.48
Q8/Q11	−0.28	.12	0.80	−2.4	0.79	−2.4	.54
Q14/Q17	−0.51	.12	0.94	−0.7	0.93	−0.8	.53
Q2/Q5	−0.61	.12	0.96	−0.4	0.95	−0.6	.44
Q8/Q9	−0.99	.12	1.31	3.5	1.28	3.2	.55
Q14/Q15	−1.20	.12	1.11	1.3	1.09	1.1	.46
Q2/Q3	−1.61	.12	1.31	3.8	1.29	3.5	.60

https://doi.org/10.33225/jbse/23.22.520

The Wright map (person-item map) visually displays the relative difficulties of items and estimates of a person's ability on the same linear logit scale (Linacre, 2020). The left-hand side of the map locates the distribution of students' ability measures, from the most capable (top) to the least capable (bottom). The right-hand side of the map presents the items from the hardest (top) to the easiest (bottom). In general, the mean of the item difficulty is centered at 0 logits. As Figure 2 shows, the 12 field items covered a range of item difficulty from −1.61 to 1.60 logits. Ordered as hypothesized, the higher the items plotted on the Wright map, the higher the complexity level, thereby providing evidence of the construct validity of the scales.

Figure 2

Wright Map for Online Metacognitive Skills



```
MEASURE     PERSON - MAP - ITEM
               <more>|<rare>
    2            .#  +T
                     |
                     |  Q17/Q18
                 ##  |  Q11/Q12
                     |
               ####  |
    1     ######### S+S Q5/Q6
                     |
          .########  |
                     |  Q7/Q10
         ##########  |  Q13/Q16
                     |  Q1/Q4
    0      .#######  +M
                     |
          .######## M|  Q8/Q11
                     |  Q14/Q17
        .##########  |  Q2/Q5
                     |
   -1     .########  +S Q8/Q9
                     |  Q14/Q15
              .####  |
                    S|
              #####  |  Q2/Q3
                     |
   -2            .#  +T
                     |
                  #  |
                     |
                  .  |
                    T|
   -3                +
                 .#  |
                     |
                     |
                  #  |
                     |
   -4                +
                     |
                     |
                 ##  |
                     |
                     |
   -5                +
               <less>|<frequent>
 EACH "#" IS 3.  EACH "." IS 1 TO 2
```

Ninth-grade Student Online Metacognitive Skills

To identify the extent of students' online metacognitive skills, this study used the mean estimates of the item difficulty of each level as the cutoff value of each complexity level. As Table 4 shows, the four cutoff values are −1.27 logits for level 1, −0.47 logits for level 2, 0.37 logits for level 3, and 1.36 logits for level 4. These values increased along with the proficiency levels as hypothesized. On average, items at higher levels are more difficult than items at lower levels.

In Rasch analysis, respondents have a 50/50 chance of correctly answering an item if they have an identical ability measure to the difficulty measure of the item (Boone et al., 2014). Therefore, students can be considered to have attained a particular level when they have the same ability measure as the cutoff value for this level. For example, a student with 0.0 logits reached Level 2 but had not yet achieved Level 3.

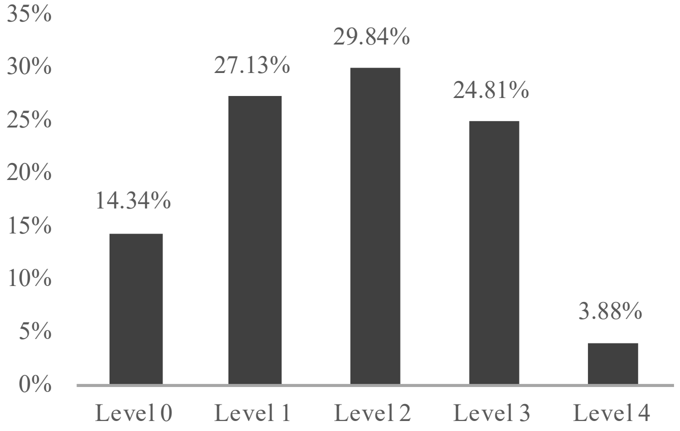
https://doi.org/10.33225/jbse/23.22.520

Table 4
Items and Measures, and Cutoff Values of Levels

Levels	Items and measures	Cutoff values
Level 1	Q2/Q3 (−1.61), Q8/Q9 (−0.99), Q14/Q15 (−1.20)	−1.27
Level 2	Q2/Q5 (−0.61), Q8/Q11 (−0.28), Q14/Q17 (−0.51)	−0.47
Level 3	Q1/Q4 (0.23), Q7/Q10 (0.51), Q13/Q16 (0.38)	0.37
Level 4	Q5/Q6 (1.03), Q11/Q12 (1.46), Q17/Q18 (1.60)	1.36

As Figure 3 shows, overall, 14.34% of ninth-grade participants reached level 0, indicating that they were not able to utilize online metacognitive skills when solving chemistry problems. Of the participants, 81.78% were classified as Level 1 (27.13%), Level 2 (29.84%), or Level 3 (24.81%), suggesting that these students were not able to accurately evaluate, regulate, or make attributions of their problem-solving performance. Only 3.88% of participants achieved Level 4, meaning that they were able to successfully employ online metacognitive skills to complete chemistry problems.

Figure 3
Online Metacognitive Skills of Sample by Level



Discussion

Following on from the paucity of research explicitly examining the online metacognitive skills of students during specific cognitive activities (Schellings, 2011; Wang, 2015), the goal of this study was to develop and validate an instrument for assessing ninth-grade students' online metacognitive skills in chemistry problem-solving. Specifically, a construct map was developed to illustrate the different levels of online metacognitive skills, and an instrument comprising 12 two-tier items was created to measure the construct. The instrument's validity and reliability were verified using Rasch analysis. Eventually, the online metacognitive performance of ninth-grade students was tested and identified according to the assessment framework. In the following section, this paper discusses the quality of the instrument and the performance of the students in online metacognitive skills.

Psychometric Properties of the Instrument

The results indicated that the instrument had good psychometric qualities. The principal components analysis confirmed that the test measures only one underlying construct. The instrument's reliability was considered acceptable, with evidence that the indices of item/person reliability and item separation were fair. The validity of the instrument was checked using fit statistics and the person-item distribution map.

https://doi.org/10.33225/jbse/23.22.520

However, there are some aspects in which the instrument could be improved. First, the person separation was not perfect, indicating the limited ability of the instrument to distinguish high and low performers. This may be because the field test sample was from the same school and grade, which might mean that it was relatively homogeneous and not widely representative. Another reason could be the inadequate items to differentiate the sample, considering that each student was administered only three tasks. Some participants did not respond to all the questions, resulting in missing data in the study, which could reduce accurate measurement (Linacre, 2020). The test is a low-stakes test and the results are not used for decision making; the relatively low person separation in this study may not be a critical issue.

The Wright map also presented several gaps in item coverage, suggesting a lack of items targeting some students. According to a widely accepted criterion, when the difficulty discrepancy between two adjacent items exceeds 0.5 logits, a gap is deemed substantively significant (Lai & Eton, 2002; Linacre, 2020). In this study, there were two obvious gaps between item Q8/Q6 (1.03 logits) and item Q7/Q10 (0.51 logits), as well as between item Q1/Q4 (0.23 logits) and item Q8/Q11 (–0.28 logits). Therefore, new items are needed to fill these gaps. Easier or harder tasks could also be designed to obtain more data about students who do exceptionally well or poorly (e.g., estimates of a person's ability under –1.61 or exceeding 1.60).

Finally, nine items had ZSTD values outside the fit range. A ZSTD of more than 2 indicates underfit, while a ZSTD of less than –2 denotes overfit (Bond & Fox, 2015). Therefore, five items showed underfit, and four items showed overfit to the Rasch model. However, for application, MNSQ is more useful than ZSTD because the former demonstrates a practical data-model fit, whereas the latter presents a perfect data-model fit (He et al., 2022). The ZSTD values can be ignored if the MNSQ values are desirable (Boone et al., 2014; Linacre, 2020). Therefore, the nine items with acceptable MNSQ are not considered problematic.

Overall, it was confirmed that the instrument assessing students' online metacognitive skills was valid and reliable.

Students' Performance in Online Metacognitive Skills

The findings of this study corroborate the work of Dermitzaki (2005) and Lawanto (2010). As the data showed, most students performed well in monitoring their performance during the problem-solving process, in line with previous literature (Azizah et al., 2019). The findings also echo Wang's (2015) study, which showed that monitoring commonly occurred among students who continually made judgments on the accuracy of their understanding and the effectiveness of their problem-solving strategies. Nearly 60% of the participants were able to evaluate their cognitive performance, similar to the findings of Wang's (2022) empirical research, which indicated that two-thirds of students reconfirmed the correctness of the solutions and checked their conclusions. However, this result differs from the findings of Overton et al. (2013), which revealed that a large portion of students had great difficulty assessing their problem-solving processes. Only one-third of the sample was able to regulate their thought processes successfully. This outcome is in agreement with other authors (Stevens et al., 2013) who noticed that many students changed their approaches to chemistry problems but did not achieve good outcomes. The ability to make attributions of their own problem-solving performance appeared to be the hardest to achieve. This may be because students rarely used effective reflective methods (Tu et al., 2020) and were confident in their results for problem-solving. Therefore, they made few causal attributions (Moller & Koller, 1999).

This study showed the four hierarchical levels of ninth graders' online metacognitive skills: monitoring, evaluating, regulating, and making attributions. This agrees with existing studies (Koriat, 2002). As Kuzle (2013) reported, students were able to monitor their understanding of the problem, but they did not assess or direct their thinking. Azizah et al. (2019) also found that students had trouble using monitoring skills and that more of them struggled with evaluation skills. Wang (2015) observed that monitoring was more common than judgment, while reflections were occasionally absent in terms of frequency of occurrence.

Overall, nearly 60% of the ninth graders in this sample were able to monitor their own thought processes or evaluate their own cognitive performance in processing chemistry problems, showing relatively good performance. This might be related to several factors. First, the field test participants were from one of the top schools in the city. In general, they have higher academic achievement than other students, which might result in better online metacognitive performance. This explanation makes sense because researchers have observed a positive correlation between academic outcomes and online metacognitive skills (Fleur et al., 2021; Treglia, 2018). Second,

https://doi.org/10.33225/jbse/23.22.520

due to cultural differences, Chinese students have high metacognitive sensitivity (van der Plas et al., 2022), which probably promotes the perception of cognitive processes.

Some students, however, did not master online metacognitive skills, and their level of skills was low. The reason may be the absence of relevant experiences (Celik, 2022), inadequate metacognitive training (Veenman, 2012), and teachers' lack of metacognitive knowledge (Heidbrink & Weinrich, 2021). Online metacognitive skills cannot be developed overnight (Pulmones, 2007). Therefore, there is an urgent need to give students more opportunities to practice online metacognitive activities, and teachers should improve their online metacognitive awareness.

Conclusions, Implications, and Limitations

In light of the absence of research to measure lower-secondary school students' online metacognitive skills, this study developed an instrument for the evaluation of ninth-grade students' online metacognitive skills in solving chemistry problems. Rasch analysis suggested that the instrument functioned as expected. The unidimensionality, fit statistics, Wright map, reliability and separation estimates verified the validity and reliability of the instrument. A sample of 258 Chinese ninth-grade students was measured and assessed using the instrument. Results showed that students' online metacognitive skills progressed along with their proficiency levels as hypothesized (i.e., monitoring, evaluating, regulating, and making attributions). However, only a few students were able to successfully employ online metacognitive skills to complete chemistry problems, while most students could not accurately monitor their cognitive processes, judge their task performance, regulate their thought processes, or make attributions of their chemistry problem-solving outcomes.

Researchers could find some inspiration from this work to generate a trustworthy measurement instrument. Guided by the Construct Modeling Approach, this study created an assessment tool using a more specific and detailed procedure that included building a construct map, generating items, developing scoring rubrics, administering a pilot test, and revising test items. Furthermore, this research may contribute to a better understanding of students' online metacognitive skills when they deal with problems and difficulties. Test designers can develop items to measure students' online metacognitive skills to offer valuable guidance and information to teachers. In addition, the instrument could serve as an available tool to enable teachers to identify students' current level of online metacognitive skills, change teaching strategies promptly, and determine the most effective educational interventions, eventually leading to better chemistry learning outcomes.

Some limitations must be considered. First, because of the limited conditions, the sample size was small, and all participants were from ninth grade in Jiangsu, China, which affected the statistical power of the analysis conducted. Second, convenience sampling may have reduced the generalizability of the results. Third, since all data were generated in chemistry problem-solving settings, the results may not generalize to other disciplines (e.g., biology, physics, earth science, engineering). Finally, this study did not conduct a think-aloud study to confirm whether participants did or did not engage in online metacognitive activities as intended.

Therefore, future research may need to include more diverse samples and more discriminative items to increase the statistical power of analyses. Further study may use a mixed-method research design to gain deeper insights and investigate the factors influencing students' online metacognitive skills. Finally, researchers could explore the connection between online metacognitive skills and chemistry problem-solving abilities, and infer whether online metacognitive activities may have a positive effect on chemistry problem-solving.

Acknowledgements

This research is supported by the Shanghai Pujiang Program (No. 2020PJC032) and the MOE Key Research Institute of Humanities and Social Sciences (No. 17JJD880007).

Conflicts of Interest

There are no conflicts to declare.

References

Azizah, U., Nasrudin, H., & Mitarlis. (2019). Metacognitive skills: A solution in chemistry problem solving. In Journal of Physics: Conference Series (Vol. 1417, p. 012084). MISEIC. https://doi.org/10.1088/1742-6596/1417/1/012084

Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning? *Metacognition and Learning, 3*(1), 39–58. https://doi.org/10.1007/s11409-007-9009-6

Bell, P., & Volckmann, D. (2011). Knowledge surveys in general chemistry: Confidence, overconfidence, and performance. *Journal of Chemical Education*, *88*(11), 1469–1476. https://doi.org/10.1021/ed100328c

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge.

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education, 90*(2), 253–269. https://doi.org/10.1002/sce.20106

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

Bryce, D., & Whitebread, D. (2012). The development of metacognitive skills: Evidence from observational analysis of young children's behavior during problem-solving. *Metacognition and Learning, 7*(3), 197–217. https://doi.org/10.1007/s11409-012-9091-2

Celik, B. (2022). The effect of metacognitive strategies on self-efficacy, motivation and academic achievement of university students. *Canadian Journal of Educational and Social Studies*, *2*(4), 37–55.

Chae, S., Park, E.-Y., & Choi, Y.-I. (2018). The psychometric properties of the Childhood Health Assessment Questionnaire (CHAQ) in children with cerebral palsy. *BMC Neurology, 18*(1), 151. https://doi.org/10.1186/s12883-018-1154-9

Chen, M. H., & Goverover, Y. (2021). Self-awareness in multiple sclerosis: Relationships with executive functions and affect. *European Journal of Neurology, 28*(5), 1627–1635. https://doi.org/10.1111/ene.14762

Cheng, C. K. E. (2011). The role of self-regulated learning in enhancing learning performance. *The International Journal of Research and Review, 6*(1), 1–16.

Clifford, M. M. (1986). The effects of ability, strategy, and effort attributions for educational, business, and athletic failure. *British Journal of Educational Psychology, 56*(2), 169–179. https://doi.org/10.1111/j.2044-8279.1986.tb02658.x

Cooper, M. M., & Sandi-Urena, S. (2009). Design and validation of an instrument to assess metacognitive skillfulness in chemistry problem solving. *Journal of Chemical Education*, *86*(2), 240–245. https://doi.org/10.1021/ed086p240

Cooper, M. M., Sandi-Urena, S., & Stevens, R. (2008). Reliable multi method assessment of metacognition use in chemistry problem solving. *Chemistry Education Research and Practice, 9*(1), 18–24. https://doi.org/10.1039/B801287N

De Clercq, A., Desoete, A., & Roeyers, H. (2000). EPA2000: A multilingual, programmable computer assessment of off-line metacognition in children with mathematical-learning disabilities. *Behavior Research Methods, Instruments, & Computers, 32*(2), 304–311. https://doi.org/10.3758/BF03207799

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

Desoete, A. (2001). *Off-line metacognition in children with mathematics learning disabilities.* Ghent University (Doctoral dissertation).

Desoete, A., & Roeyers, H. (2002). Off-line metacognition — A domain-specific retardation in young children with learning disabilities? *Learning Disability Quarterly, 25*(2), 123–139. https://doi.org/10.2307/1511279

Efklides, A. (2002). The systemic nature of metacognitive experiences. In P. Chambres, M. Izaute & P. J. Marescaux (Eds.), *Metacognition: Process, function and use* (pp. 19–34). Springer. https://doi.org/10.1007/978-1-4615-1099-4_2

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist, 13*(4), 277–287. https://doi.org/10.1027/1016-9040.13.4.277

Efklides, A., & Misailidi, P. (2010). Introduction: The present and the future in metacognition. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 1–18). Springer. https://doi.org/10.1007/978-1-4419-6546-2_1

Efklides, A., Papadaki, M., Papantoniou, G., & Kiosseoglou, G. (1998). Individual differences in feelings of difficulty: The case of school mathematics. *European Journal of Psychology of Education, 13*(2), 207–226. https://doi.org/10.1007/BF03173090

Engelhard, G., & Myford, C. M. (2003). *Monitoring faculty consultant performance in the advanced placement English Literature and Composition program with a many-faceted Rasch model*. New York: College Entrance Examination Board.

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions, 21*(1), 1095.

Fitzpatrick, R., Norquist, J. M., Jenkinson, C., Reeves, B. C., Morris, R. W., Murray, D. W., & Gregg, P. J. (2004). A comparison of Rasch with Likert scoring to discriminate between patients' evaluations of total hip replacement surgery. *Quality of Life Research, 13*(2), 331–338. https://doi.org/10.1023/B:QURE.0000018489.25151.e1

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fleur, D. S., Bredeweg, B., & van den Bos, W. (2021). Metacognition: Ideas and insights from neuro- and educational sciences. *NPJ Science of Learning*, *6*. https://doi.org/10.1038/s41539-021-00089-5

Gamby, S., & Bauer, C. F. (2022). Beyond "study skills": A curriculum-embedded framework for metacognitive development in a college chemistry course. *International Journal of STEM Education*, *9*(1), 61–61. https://doi.org/10.1186/s40594-022-00376-6

Garner, J. K. (2009). Conceptualizing the relations between executive functions and self-regulated learning. *The Journal of Psychology, 143*(4), 405–426. https://doi.org/10.3200/JRLP.143.4.405-426

Gilbert, J. K. (2005). Visualization: A metacognitive skill in science and science education. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 9–27). Springer. https://doi.org/10.1007/1-4020-3613-2_2

Goldstein, S., & Naglieri, J. A. (2011). *Encyclopedia of child behavior and development*. Springer.

Hawker, M. J., Dysleski, L., & Rickey, D. (2016). Investigating general chemistry students' metacognitive monitoring of their exam performance by measuring postdiction accuracies over time. *Journal of Chemical Education*, *93*(5), 832–840. https://doi.org/10.1021/acs.jchemed.5b00705

He, P., Zheng, C., & Li, T. (2022). Upper secondary school students' conceptions of chemical equilibrium in aqueous solutions: Development and validation of a two-tier diagnostic instrument. *Journal of Baltic Science Education*, *21*(3), 428–444. https://doi.org/10.33225/jbse/22.21.428

Heidbrink, A., & Weinrich, M. (2021). Undergraduate chemistry instructors' perspectives on their students' metacognitive development. *Chemistry Education Research and Practice, 22*(1), 182–198. https://doi.org/10.1039/D0RP00136H

Hollingworth, R. W., & McLoughlin, C. (2001). Developing science students' metacognitive problem solving skills online. *Australasian Journal of Educational Technology, 17*(1), 50–63. https://doi.org/10.14742/ajet.1772

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

Kinnunen, R., & Vauras, M. (2010). Tracking on-line metacognition: Monitoring and regulating comprehension in reading. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 209–229). Springer. https://doi.org/10.1007/978-1-4419-6546-2_10

Kipnis, M., & Hofstein, A. (2008). The inquiry laboratory as a source for development of metacognitive skills. *International Journal of Science and Mathematics Education, 6*(3), 601–627. https://doi.org/10.1007/s10763-007-9066-y

Koren, D., Poyurovsky, M., Seidman, L. J., Goldsmith, M., Wenger, S., & Klein, E. M. (2005). The neuropsychological basis of competence to consent in first-episode schizophrenia: A pilot metacognitive study. *Biological Psychiatry, 57*(6), 609–616. https://doi.org/10.1016/j.biopsych.2004.11.029

Koriat, A. (2002). Metacognition research: An interim report. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 261–286). Cambridge University Press. https://doi.org/10.1017/CBO9780511489976.012

Kuzle, A. (2013). Patterns of metacognitive behavior during mathematics problem-solving in a dynamic geometry environment. *International Electronic Journal of Mathematics Education, 8*(1), 20–40. https://doi.org/10.29333/iejme/272

Kuzle, A. (2018). Assessing metacognition of grade 2 and grade 4 students using an adaptation of multi-method interview approach during mathematics problem-solving. *Mathematics Education Research Journal, 30*(2), 185–207. https://doi.org/10.1007/s13394-017-0227-1

Lai, J.-S., & Eton, D. T. (2002). Clinically meaningful gaps. *Rasch Measurement Transactions*, *15*(4), 850.

Lavi, R., Shwartz, G., & Dori, Y. J. (2019). Metacognition in chemistry education: A literature review. *Israel Journal of Chemistry, 59*, 583–597. https://doi.org/10.1002/ijch.201800087

Lawanto, O. (2010). Students' metacognition during an engineering design project. *Performance Improvement Quarterly, 23*(2), 117–136. https://doi.org/10.1002/piq.20084

Lee, N. H., Ng, K. E. D., & Yeo, J. B. W. (2019). Metacognition in the teaching and learning of mathematics. In T. L. Toh, B. Kaur & E. G. Tay (Eds.), *Mathematics education in Singapore* (pp. 241–268). Springer. https://doi.org/10.1007/978-981-13-3573-0_11

Lin, C.-H. (2003). *Intergenerational parallelism of self-efficacy: Moderating variables, mediating variables, and common antecedents.* Texas A&M University (Doctoral dissertation).

Linacre, J. M. (2009). *A user's guide to WINSTEPS: Rasch-model computer programs: program manual 3.72.3*. Mesa-Press.

Linacre, J. M. (2020). *A user's guide to WINSTEPS/MINISTEP: Rasch-model computer programs*. Winsteps.com.

Liu, X. F. (2020). *Using and developing measurement instruments in science education: a Rasch modeling approach*. IAP.

Mathabathe, K. C., & Potgieter, M. (2017). Manifestations of metacognitive activity during the collaborative planning of chemistry practical investigations. *International Journal of Science Education*, *39*(11), 1465–1484. https://doi.org/10.1080/09500693.2017.1336808

Mazancieux, A., Moulin, C. J. A., Casez, O., & Souchay, C. (2021). A multidimensional assessment of metacognition across domains in multiple sclerosis. *Journal of the International Neuropsychological Society, 27*(2), 124–135. https://doi.org/10.1017/S1355617720000776

McCord, R. E., & Matusovich, H. M. (2019). Naturalistic observations of metacognition in engineering: Using observational methods to study metacognitive engagement in engineering. *Journal of Engineering Education, 108*(4), 481–502. https://doi.org/10.1002/jee.20291

Moller, J., & Koller, O. (1999). Spontaneous cognitions following academic test results. *The Journal of Experimental Education*, *67*(2), 150–164. https://doi.org/10.1080/00220979909598350

Muller, P. C. (2005). *Examining the psychometric properties of the School Violence Inventory using item response theory.* University of California (Doctoral dissertation).

Ng, K. E. D., Lee, N. H., & Safii, L. (2021). Facilitation of students' metacognition: Some insights gleaned from mathematics classrooms in Singapore secondary schools. In B. Kaur & Y. H. Leong (Eds.), *Mathematics instructional practices in Singapore secondary schools* (pp. 105–122). Springer. https://doi.org/10.1007/978-981-15-8956-0_6

Overton, T., Potter, N., & Leng, C. (2013). A study of approaches to solving open-ended problems in chemistry. *Chemistry Education Research and Practice, 14*(4), 468–475. https://doi.org/10.1039/C3RP00028A

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research, 15*(2), 020111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Pulmones, R. (2007). Learning chemistry in a metacognitive environment. *The Asia-Pacific Education Researcher, 16*(2), 165–183. https://doi.org/10.3860/taper.v16i2.258

Quiles, C., Prouteau, A., & Verdoux, H. (2020). Assessing metacognition during or after basic-level and high-level cognitive tasks? A comparative study in a non-clinical sample. *L'Encéphale, 46*(1), 3–6. https://doi.org/10.1016/j.encep.2019.05.007

Quiles, C., Verdoux, H., & Prouteau, A. (2014). Assessing metacognition during a cognitive task: Impact of "on-line" metacognitive questions on neuropsychological performances in a non-clinical sample. *Journal of the International Neuropsychological Society, 20*(5), 547–554. https://doi.org/10.1017/S1355617714000290

Rickey, D., & Stacy, A. M. (2000). The role of metacognition in learning chemistry. *Journal of Chemical Education, 77*(7), 915. https://doi.org/10.1021/ed077p915

Rittle-Johnson, B., Matthews, P. G., Taylor, R. S., & McEldoon, K. L. (2011). Assessing knowledge of mathematical equivalence: A construct-modeling approach. *Journal of Educational Psychology, 103*(1), 85–104. https://doi.org/10.1037/a0021334

Ross, J. D. (1999). *Regulating hypermedia: Self-regulation learning strategies in a hypermedia environment.* Virginia Polytechnic Institute and State University (Doctoral dissertation).

Rozencwajg, P. (2003). Metacognitive factors in scientific problem-solving strategies. *European Journal of Psychology of Education, 18*(3), 281–294. https://doi.org/10.1007/bf03173249

Sandi-Urena, S., Cooper, M. M., & Stevens, R. H. (2011). Enhancement of metacognition use and awareness by means of a collaborative intervention. *International Journal of Science Education, 33*(3), 323–340. https://doi.org/10.1080/09500690903452922

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. 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*(3), 963–990. https://doi.org/10.1007/s10212-012-0149-y

Seel, N. M. (2012). Metacognition and learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2228–2231). Springer. https://doi.org/10.1007/978-1-4419-1428-6_108

She, H. C., Cheng, M. T., Li, T. W., Wang, C.-Y., Chiu, H. T., Lee, P. Z., Chou, W. C., & Chuang, M. H. (2012). Web-based undergraduate chemistry problem-solving: The interplay of task performance, domain knowledge and web-searching strategies. *Computers & Education, 59*(2), 750–761. https://doi.org/10.1016/j.compedu.2012.02.005

Sideridis, G. D. (2007). Persistence of performance-approach individuals in achievement situations: An application of the Rasch model. *Educational Psychology, 27*(6), 753–770. https://doi.org/10.1080/01443410701309290

Smith, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling, 3*(1), 25–40. https://doi.org/10.1080/10705519609540027

Soto, D., Theodoraki, M., & Paz-Alonso, P. M. (2018). How the human brain introspects about one's own episodes of cognitive control. *Cortex, 107*, 110–120. https://doi.org/10.1016/j.cortex.2017.10.016

Sporer, S. L., & Horry, R. (2011). Pictorial versus structural representations of ingroup and outgroup faces. *Journal of Cognitive Psychology, 23*(8), 974–984. https://doi.org/10.1080/20445911.2011.594434

Stevens, R., Beal, C. R., & Sprang, M. (2013). Assessing students' problem solving ability and cognitive regulation with learning trajectories. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 409–423). Springer. https://doi.org/10.1007/978-1-4419-5546-3_27

Teichert, M. A., Tien, L. T., Dysleski, L., & Rickey, D. (2017). Thinking processes associated with undergraduate chemistry students' success at applying a molecular-level model in a new context. *Journal of Chemical Education, 94*(9), 1195–1208. https://doi.org/10.1021/acs.jchemed.6b00762

Testa, I., Galano, S., & Tarallo, O. (2023). The relationships between freshmen's accuracy of self-evaluation and the likelihood of succeeding in chemistry and physics exams in two STEM undergraduate courses. *International Journal of Science Education, 45*(5), 358–382. https://doi.org/10.1080/09500693.2022.2162833

Treglia, M. (2018). *A Comparison of Offline and Online Measures of Metacognition.* Trinity College (Bachelor dissertation).

Tu, T., Li, C., Zhou, Z., & Guo, G. (2020). Students' difficulties with partial differential equations in quantum mechanics. *Physical Review Physics Education Research*, *16*(2), 020163. https://doi.org/10.1103/PhysRevPhysEducRes.16.020163

van der Plas, E., Zhang, S., Dong, K., Bang, D., Li, J., Wright, N. D., & Fleming, S. M. (2022). Identifying cultural differences in metacognition. *Journal of Experimental Psychology. General*, *151*(12), 3268–3280. https://doi.org/10.1037/xge0001209

Veenman, M. V. J. (2012). Metacognition in science education: Definitions, constituents, and their intricate relation with cognition. In A. Zohar & Y. J. Dori (Eds.), *Metacognition in science education* (pp. 21–36). Springer. https://doi.org/10.1007/978-94-007-2132-6_2

Veenman, M. V. J., van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Veenman, M. V. J. (2013). Assessing metacognitive skills in computerized learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 157–168). Springer. https://doi.org/10.1007/978-1-4419-5546-3_11

Veenman, M. V. J. (2017). Assessing metacognitive deficiencies and effectively instructing metacognitive skills. *Teachers College Record, 119*(13), 1–20. https://doi.org/10.1177/016146811711901303

Veenman, M. V. J., & van Cleef, D. (2019). Measuring metacognitive skills for mathematics: Students' self-reports versus on-line assessment methods. *ZDM, 51*(4), 691–701. https://doi.org/10.1007/s11858-018-1006-5

Violeau, L., Dudilot, A., Roux, S., & Prouteau, A. (2020). How internalised stigma reduces self-esteem in schizophrenia: The crucial role of off-line metacognition. *Cognitive Neuropsychiatry, 25*(2), 154–161. https://doi.org/10.1080/13546805.2020.1714570

Vo, K., Sarkar, M., White, P. J., & Yuriev, E. (2022). Problem solving in chemistry supported by metacognitive scaffolding: Teaching associates' perspectives and practices. *Chemistry Education Research and Practice, 23*(2), 436–451. https://doi.org/10.1039/D1RP00242B

Wang, C.-Y (2022). Evaluating the effects of the analogical learning approach on eighth graders' learning outcomes: The role of metacognition. *Chemistry Education Research and Practice, 24*(2), 535–550. https://doi.org/10.1039/d2rp00074a

Wang, C.-Y. (2015). Exploring general versus task-specific assessments of metacognition in university chemistry students: A multitrait–multimethod analysis. *Research in Science Education, 45*(4), 555–579. https://doi.org/10.1007/s11165-014-9436-8

Wang, Z., Chi, S., Luo, M., Yang, Y., & Huang, M. (2017). Development of an instrument to evaluate high school students' chemical symbol representation abilities. *Chemistry Education Research and Practice, 18*(4), 875–892. https://doi.org/10.1039/C7RP00079K

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., Dolan, R. J., & Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition, 22*(1), 264–271. https://doi.org/10.1016/j.concog.2013.01.004

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410611697

Wilson, M., Allen, D. D., & Li, J. C. (2006). Improving measurement in health education and health behavior research using item response modeling: Introducing item response modeling. *Health Education Research, 21*, i4–i18. https://doi.org/10.1093/her/cyl108

Wind, S. A., & Schumacker, R. E. (2021). Exploring the impact of missing data on residual-based dimensionality analysis for measurement models. *Educational and Psychological Measurement, 81*(2), 290–318. https://doi.org/10.1177/0013164420939634

Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning, 9*(2), 229–237. https://doi.org/10.1007/s11409-014-9113-3

Wong, K. Y. (2007). *Metacognitive awareness of problem solving among primary and secondary school students*. Paper presented at the Redesigning Pedagogy: Culture, Knowledge and Understanding Conference.

Wren, D., & Barbera, J. (2014). Psychometric analysis of the thermochemistry concept inventory. *Chemistry Education Research and Practice, 15*(3), 380–390. https://doi.org/10.1039/C3RP00170A

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist, 48*(3), 135–147. https://doi.org/10.1080/00461520.2013.794676

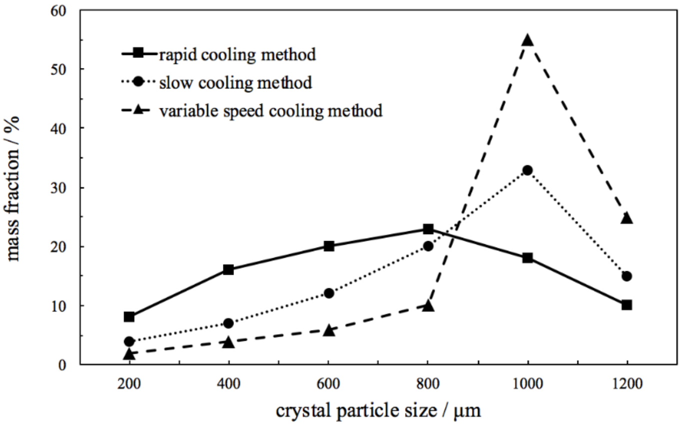
Zimmerman, B. J. (2011). Motivational sources and outcomes of self-regulated learning and performance. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 63–78). Routledge. https://doi.org/10.4324/9780203839010

Zoller, U., & Pushkin, D. (2007). Matching higher-order cognitive skills (HOCS) promotion goals with problem-based laboratory practice in a freshman organic chemistry course. *Chemistry Education Research and Practice, 8*(2), 153–171. https://doi.org/10.1039/B6RP90028C

Appendix A. An example task

CRYSTALLIZATION

Potassium nitrate partially precipitates in the form of crystals when cooling a hot concentrated aqueous solution of potassium nitrate. There are many cooling methods, such as rapid cooling, slow cooling, and variable speed cooling. Different cooling rates affect the degree of crystallization. A student used three cooling methods to crystallize the potassium nitrate, and the crystal particle size distribution is shown in the figure below. Which cooling method should be chosen to obtain large crystals with relatively uniform particle sizes? (Do not answer this question here. Please answer the following questions in sequence.)



Q13. Before answering the above question, please estimate the probability that you can answer the question correctly. Tick "√" in the table below according to your judgment.

incorrect correct

0%	25%	50%	75%	100%

Q14. According to the figure above, which cooling method should be chosen to obtain large crystals with relatively uniform particle sizes? Please explain the thinking processes and ideas you used to come to your answer.

Q15. Have you recognized if the thinking processes are clear and whether the reasoning is correct or incorrect when solving this chemistry problem?
A. Yes. My thinking is clear and my reasoning is correct.
B. Yes. My thinking is not clear, and I encounter difficulties in reasoning about this problem.
C. No. I do not have such recognition.

Q16. Based on your answers in Q14, please re-estimate the probability that you can answer this question correctly. Tick "√" in the table below according to your judgment.

incorrect correct

0%	25%	50%	75%	100%

Q17. Please evaluate the answer you gave to Q14?
A. My answer is correct.
B. My answer is partially correct.
C. My answer is incorrect.
D. It is hard for me to evaluate my answer.

Q18. Why did you make the evaluation in Q17?

Item Q14 and item Q15 (abbreviated to Q14/Q15) were designed in accordance with Level 1, which assessed whether students monitored their own thinking during the process of solving a chemistry problem. The first tier required students to answer a question about which cooling method should be chosen to obtain large crystals with relatively uniform particle sizes and to write down their thinking processes. The second tier contained three options that depicted different monitoring awareness about the thinking processes when solving the first-tier question and required students to choose one that best fitted their situation. The response demands of Q14 and Q17 (abbreviated to Q14/Q17) were in line with Level 2, which assessed online metacognitive evaluation. The second tier consisted of four options from which students had to choose one to indicate their evaluation results on the answers they gave in the first tier. Q13 and Q16 (abbreviated to Q13/Q16) were generated to measure online metacognitive regulation (Level 3). Immediately after reading the scenario and question, students were required to estimate the likelihood of giving an accurate response to the question in the first tier. After writing down their answers, respondents were asked to rate the probability again in the second tier. The difference between the two probability values was used to infer whether students might have regulated their thinking and thus changed their awareness about whether the problem could be solved. The Q17 and Q18 (abbreviated to Q17/Q18) belonged to Level 4, examining the ability of students to make attributions of their own chemistry problem-solving performance. Students evaluated their problem-solving performance in the first tier, and then they were provided with a space to write down the reasons for their evaluation in the second tier.

Appendix B. The scoring rubrics

Level	Items	Score	Performance	Evaluation criteria
Level 1	Q2/Q3 Q8/Q9 Q14/Q15 Q20/Q21 Q26/Q27	2	Participants can accurately monitor their own thought processes.	The first-tier responses are correct, and the second-tier responses are option A. The first-tier responses are partially correct, and the second-tier responses are option B. The first-tier responses are incorrect, and the second-tier responses are option B.
		1	Participants can monitor their own thought processes, but not accurately enough.	The first-tier responses are correct, and the second-tier responses are option B. The first-tier responses are partially correct, and the second-tier responses are option A. The first-tier responses are incorrect, and the second-tier responses are option A.
		0	Participants cannot monitor their own thought processes.	The second-tier responses are option C.
Level 2	Q2/Q5 Q8/Q11 Q14/Q17 Q20/Q23 Q26/Q29	2	Participants can accurately evaluate their own cognitive performance.	The first-tier responses are correct, and the second-tier responses are option A. The first-tier responses are partially correct, and the second-tier responses are option B. The first-tier responses are incorrect, and the second-tier responses are option C.
		1	Participants can evaluate their own cognitive performance, but not accurately enough.	The first-tier responses are correct, and the second-tier responses are option B. The first-tier responses are correct, and the second-tier responses are option C. The first-tier responses are partially correct, and the second-tier responses are option A. The first-tier responses are partially correct, and the second-tier responses are option C. The first-tier responses are incorrect, and the second-tier responses are option A. The first-tier responses are incorrect, and the second-tier responses are option B.
		0	Participants cannot evaluate their own cognitive performance.	The second-tier responses are option D.

https://doi.org/10.33225/jbse/23.22.520

Level	Items	Score	Performance	Evaluation criteria
Level 3	Q1/Q4 Q7/Q10 Q13/Q16 Q19/Q22 Q25/Q28	2	Participants can successfully regulate their own thought processes, which makes their problem-solving performance better.	The probability rated by participants in the second tier is higher than the probability rated in the first tier.
		1	Participants cannot successfully regulate their own thought processes, maintaining the same problem-solving performance.	The probability rated by participants in the second tier is equal to the probability rated in the first tier.
		0	Participants cannot regulate their own thought processes, which makes their problem-solving performance worse.	The probability rated by participants in the second tier is lower than the probability rated in the first tier.
Level 4	Q5/Q6 Q11/Q12 Q17/Q18 Q23/Q24 Q29/Q30	1	Participants can make attributions of their own cognitive performance.	Reasonable explanations for the problem-solving performance are given in the second tier.
		0	Participants cannot make attributions of their own cognitive performance.	Unreasonable explanations for the problem-solving performance are given in the second tier.

Received: *September 09, 2022* Revised: *February 24, 2023* Accepted: *April 16, 2023*

Cite as: Zeng, Y., Chi, S., Wang, Z., & Zhuang, X. (2023). Developing and validating an instrument to assess ninth-grade students' online metacognitive skills in solving chemistry problems. *Journal of Baltic Science Education*, *22*(3), 520-537. https://doi.org/10.33225/jbse/23.22.520

Yating Zeng
MEd, (Science Education), PhD Candidate, College of Teacher Education, Faculty of Education, East China Normal University, Shanghai, China.
E-mail: yatingzeng1003@163.com
ORCID: https://orcid.org/0000-0001-9003-0382

Shaohui Chi
(Corresponding author)
PhD, Associate Professor, College of Teacher Education, Faculty of Education, East China Normal University, Shanghai, China.
E-mail: charlcy@163.com
ORCID: https://orcid.org/0000-0002-1261-7951

Zuhao Wang
PhD, Professor, College of Teacher Education, Faculty of Education, East China Normal University, Shanghai, China.
E-mail: wangzuhao@126.com
ORCID: https://orcid.org/0000-0002-2967-8232

Xiaosong Zhuang
PhD, College of Teacher Education, Faculty of Education, East China Normal University, Shanghai, China.
E-mail: zhuangxiaosongfeel@126.com