Assessing Metacognitive Skills in Mathematical Problem Solving: A Comparative Study of a Task-Specific Questionnaire and Think Aloud Protocol

Uthpala Athukorala, Institute of Technology University of Moratuwa, Sri Lanka
Dileepa Fernando, Singapore University of Technology and Design, Singapore
Chanakya Wijeratne, University of Colombo, Sri Lanka

The European Conference on Education 2024
Official Conference Proceedings

Abstract
Metacognitive skills play a crucial role in aiding students to comprehend when, why, where, and how to apply their knowledge to mathematical problem-solving. Consequently, teachers seek effective measurements to assess metacognitive skills and identify students with low metacognitive skills in mathematical problem-solving. Think Aloud Protocol (TAP) is recommended as the best method for measuring metacognitive skills in Mathematics. But using TAP is impractical for classes with a large number of students due to administrative challenges and time constraints. Questionnaires, although have less administrative issues, they are not generating reliable data comparable to TAP. To address these challenges, a task-specific questionnaire was designed based on the TAP, boasting a reliability of 0.89 (ordinal alpha) and a content validation score (S-CVI/Ave) of 0.9. The primary objective of this study is to examine whether the data generated by the questionnaire align through data generated by TAP. Twenty (20), third-year undergraduates from the Sri Lanka Technological Campus (SLTC) pursuing a BSc in Information Technology participated in the study. Engaging in the problem-solving process, they verbalized their thoughts, and immediately filled out the newly designed questionnaire. The overall correlation between the questionnaire and TAP (r=0.52) suggests the potential utility of this questionnaire as an alternative to TAP.

Keywords: Assessing Metacognitive Skills, Mathematics, Questionnaire, Thinking Aloud Protocol

1. Introduction

The Covid pandemic in 2020 brought significant transformations to the education sector. With students and teachers confined to their homes to prevent the virus's spread, traditional in-person classes became impossible. However, the passage of time and the need to continue education compelled schools, colleges, and universities to adapt swiftly. They transitioned from physical classrooms to virtual ones. Utilizing online meeting platforms like Zoom, educators worldwide resumed their lessons. This shift marked a profound change in the education landscape over the past five years.

Students remained at home, utilizing online meeting platforms to interact with their teachers and Learning Management Systems (LMS) for accessing notes and completing assignments. This shift eliminated direct connections with teachers and peers, requiring students to manage their studies independently. Students needed a skill to set learning goals and monitor, regulate, and control their cognition, motivation, and behavior according to these goals (Artino, 2007), a skill known as "Self-Regulated Learning (SRL)."

Improving self-regulated learning skills in students, specially who learn in online platforms is essential where teacher interaction and peer support is less (Dabbagh & Kitsantas, 2005). Metacognition is one of the three major components of SRL which describes students' ability to understand and control their cognitive process (Schraw et al., 2006). Having an idea about the students' metacognitive level supports the teacher to take necessary precautions. Hence assessing metacognitive level of the students in the classroom make the opportunity for teachers to expand their support for the students in a broad perspective.

Assessing metacognition is challenging because it is a complex structure that is not externally visible. It is sometimes affected by verbal ability and working memory capacity (Akturk & Sahin, 2011). Previous studies recommend that online measures are preferred for assessing metacognitive skills in mathematical problem solving (Veenman & van Cleef, 2019). However, practical issues with using online measures, especially in large classes, discourage teachers from using them. If a validated questionnaire that has a positive relationship with an online measure and generates accurate responses can replace an online measure with practical issues, it would resolve the usability concerns of online measures. This study aims to examine the relationship between the Questionnaire on Measuring Metacognitive Skills in Mathematical Problem Solving (Athukorala et al., 2024) and the Think Aloud Protocol (TAP) to determine the compatibility of the questionnaire with TAP.

Next sections describe the available methods for measuring metacognition, methodology followed, results and discussion.

2. Literature Review

Research has explored the relationship between SRL and online learning. Dabbagh and Kitsantas (2005) argued that students in digital environments require higher levels of self-regulatory skills compared to those in physical classrooms, due to reduced instructor support and less monitoring of students' attention and progress. An early study by Kinzie (1990) identified self-regulatory skills as one of the three critical factors influencing students' success in online environments. The effective use of SRL is crucial in digital learning settings where students have a high degree of autonomy due to the absence of a physically present instructor (Kinzie, 1990).

Previous studies indicate that learners often struggle with online learning due to a lack of critical self-regulated learning (SRL) skills (Azevedo, 2005). Further research supports the idea that SRL is essential for successful online learning. Broadbent and Poon (2015) discovered a positive and significant relationship between SRL and success in online learning, reinforcing this notion. Therefore, providing SRL support for online learners is likely to enhance their academic success in virtual environments (Wong et al., 2019).

According to Zimmerman, taking responsibility for their own learning involves students being metacognitively, motivationally, and behaviorally engaged in the learning process (Wong et al., 2019). In traditional learning environments, students with effective self-regulated learning (SRL) skills are recognized by researchers and educators as the most successful learners (Wong et al., 2019). Similarly, SRL skills are even more crucial in online environments, which feature high levels of learner autonomy and minimal teacher presence (Wong et al., 2019).

Self-regulation comprises three main components: cognition, metacognition, and motivation (Figure 1) (Schraw et al., 2006). Cognition involves the ability to encode, memorize, and recall information. Metacognition refers to students' capacity to understand and monitor their cognitive processes. Motivation encompasses the beliefs and attitudes that influence the application and development of cognitive and metacognitive abilities. While these three elements are essential for self-regulation, they are not sufficient on their own (Schraw et al., 2006).

Broadbent and Poon (2015) identified four self-regulated learning strategies that significantly impact online learning: metacognition, time management, effort regulation, and critical thinking. Their study highlights that metacognition is a critical factor in developing effective self-regulation strategies in digital learning environments.

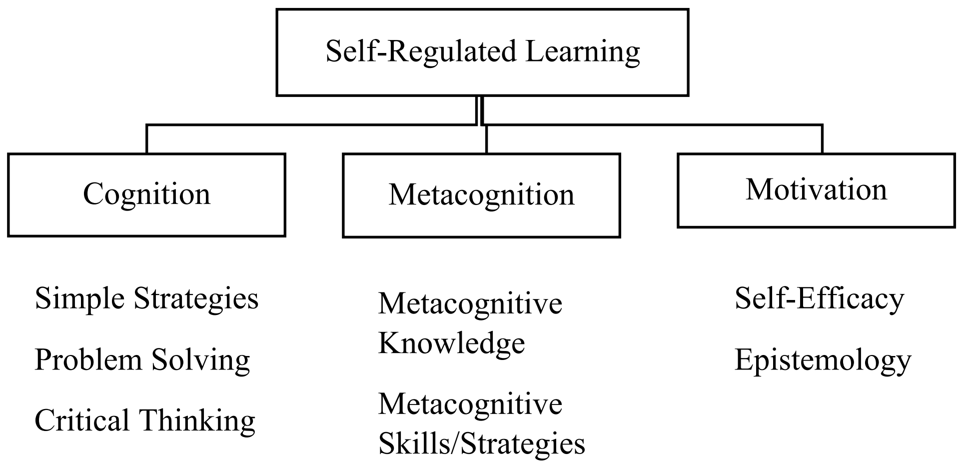

Figure 1: Components of SRL

Metacognition, simply put, is "thinking about thinking" (Akturk & Sahin, 2011). It involves reflecting on and regulating one’s own mental processes. Metacognition is the process of monitoring and evaluating comprehension during learning, representing a higher level of thinking about how to approach and manage a learning task.

Metacognition consists of two main components (Figure 2): metacognitive knowledge and metacognitive skills (Schraw et al., 2006). Metacognitive knowledge includes:

1) Declarative Knowledge: Understanding about the learner themselves.
2) Procedural Knowledge: Knowing strategies.
3) Conditional Knowledge: Understanding why and when to use a particular strategy in a specific context (Schraw et al., 2006).

This type of knowledge involves tasks like taking notes and summarizing main ideas, which are specific to the task at hand (Schuster et al., 2020).

Metacognitive skills, also known as metacognitive strategies, include three components based on Flavell's original categorization:

1) Planning: Deciding how to approach a learning task.
2) Monitoring: Keeping track of one's comprehension and performance.
3) Evaluation: Assessing the effectiveness of one's learning strategies and processes.

Within Self-Regulated Learning (SRL), cognition, motivation, and metacognitive knowledge are non-transferable and task-specific components that directly address specific learning tasks. Conversely, metacognitive skills, which manage a broad range of cognitive and motivational tasks, are task-general and transferable. Therefore, enhancing metacognitive skills is particularly beneficial, as these skills can be applied across various learning contexts.

At the same time, enhancing metacognitive skills can contribute to improving Self-Regulated Learning (SRL) strategies among online learners, as metacognitive skills are a major component of SRL. Hence, assessing metacognitive skills of the students can support to identify who need the improvements.

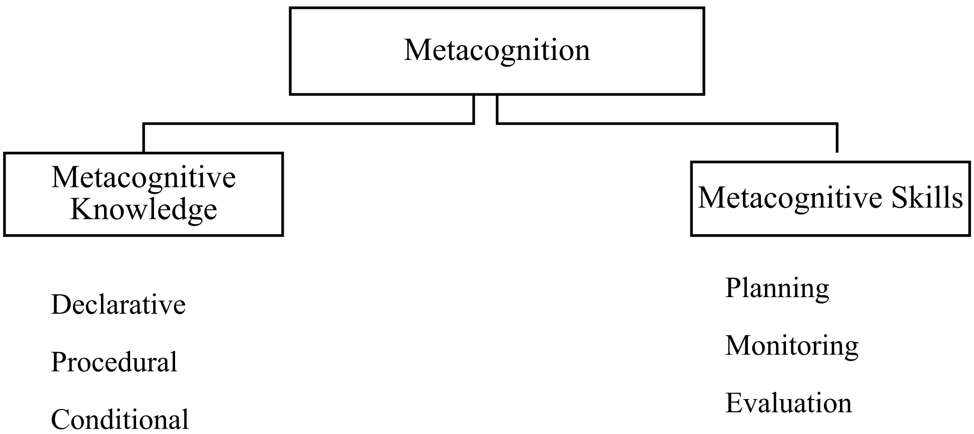

Figure 2: Components of Metacognition

2.1 Available Methods for Measuring Metacognition

For measuring metacognition, various methods can assess metacognitive knowledge and skills either separately or together. According to available literature, these measurement methods can be applied in different settings, categorized as follows:

i. **Method of Implementation**: Offline or Online
ii. **Behavior**: Objective behavior or Self-reporting
iii. **Applied Environment**: Physical classroom or Digital learning environment

i. Method of Implementation

Measurements of metacognition can indeed be categorized into offline and online measures based on when they are administered relative to the task:

a. **Online Measures**: These are taken while the student is actively performing a task. Examples include: Think-aloud protocol, Systematic observations
b. **Offline Measures**: These are administered either before (prospective) or after (retrospective) the task: Interviews, Questionnaires

These categorizations help researchers and educators choose appropriate methods to assess different aspects of metacognition in various educational settings.

ii. Behavior

Objective behavioral measurements in measuring metacognitive skills involve assessing observable behaviors or actions that reflect an individual's metacognitive processes. These measurements emphasize direct observation rather than relying solely on self-reporting (Ozturk, 2017). Examples of objective behavioral measurements include: think aloud protocol, systematical observations, eye tracking in digital learning environment, response time and error tracking. These objective behavioral measures are valuable for assessing how individuals apply metacognitive skills in real-time learning situations, offering a deeper understanding of their cognitive and metacognitive functioning beyond self-reported data.

iii. Environment

There are measures where we can apply physical classrooms and the digital learning environment. Interviews, questionnaires, think aloud protocol, and systematical observations can be applied in physical classroom settings. Online self-report questionnaires, eye tracking, log data analytics, or interactive learning analytics dashboards are some examples that can be applied in the digital environment (Ozturk, 2017; Veenman, 2013; Veenman et al., 2014). Most popular methods that were identified are questionnaires, interviews, and thinking aloud protocol (Ozturk, 2017). Log file analysis and eye tracking in the digital environment are trending. Some common measurement types are discussed below.

2.2 Common Measurement Types

i. Questionnaires

In questionnaires, students present their actions related to metacognition as answers to questions. Sometimes they may interpret their own metacognitive behavior or demonstrate their understanding about metacognition or apply metacognitive strategies. Some popular questionnaires measuring metacognition are tabled below.

Name	Metacognitive Knowledge /Skills or both	Description
1. Metacognitive Awareness of Reading Strategies Inventory (MARSI)	Metacognitive knowledge	To asses adolescent and adult readers' reading strategies and problem solving strategies (Ozturk, 2017).
2. Metacognitive Awareness Inventory (MAI)	Both	To assess adult's general metacognitive knowledge and skills (Ozturk, 2017).
3. Learning and Study Strategy Inventory (LASSI)	Metacognitive skills	Likert type questionnaire on individual's domain general and domain specific knowledge of metacognitive skills/regulation (Ozturk, 2017).
4. Motivated Strategies for Learning Questionnaire (MSLQ)	Metacognitive skills	This is a Likert type questionnaire which assess rehearsal, elaboration, organization, critical thinking, metacognitive monitoring and self-regulated learning.
5. Survey of Reading Strategies (SORS)	Metacognitive skills	Intends to measure use of strategies while reading academic materials.
6. Awareness of Independent Learning Inventory (AILI)	Metacognitive Knowledge, Regulation and Responsiveness.	Valid and reliable instrument for assessing effects of interventions

Table 1: Questionnaires for Measuring Metacognition

ii. Interviews

In interviews, students are asked questions about their metacognitive behavior, self-regulated learning strategies etc. Based on the answers given students' level of metacognition is evaluated. One common interview method is named as Self-Regulated Learning Interview Schedule (SRLIS). Students are asked questions on self-regulation and after that they are asked how they would behave in different situations like classroom discussion, short writing assignment, mathematics assignments, end of term tests, studying at home etc. Their responses are categorized into knowledge, metacognitive strategy use and regulation (Ozturk, 2017).

iii. Think Aloud Protocol

In think aloud protocol, participants are advised to verbalize their thoughts while performing their tasks. Task may be a laboratory experiment, mathematical problem solving or reading exercise. Only in case, participants are silent, assessor may tell "keep talking". Think aloud processes are transcribed and assessed using coding system (Ozturk, 2017). By those coding systems evaluators can identify and quantify the level of metacognition strategy usage, regulation and knowledge (Ozturk, 2017; Veenman & van Cleef, 2019). At least there should be two raters to assess think aloud activity. After evaluation, both raters get together and discuss about their evaluation and come to a conclusion at points where they have given different marks (Jacobse & Harskamp, 2012; Ozturk, 2017). Think aloud is the only method that opens the opportunity for accessing students' thoughts and metacognitive deliberations parallel to task performance (Veenman & van Cleef, 2019).

iv. Systematical Observations

While student is performing the task, judges observe the activities done by the participants. Judges do not disturb the participants. Often used with young children. As in the think aloud method, there should be a coding system which includes all the possible activities for the evaluation (Ozturk, 2017).

v. Self-Report Judgement

Students are asked to rank their leaning ability before the task assignment. Then, after completing the given task, students again rank their learning and give judgement of their learning. Each student's actual learning is compared with their confidence in learning. This is metacognition monitoring. Both the students who felt they know and do and the students who felt they don't know and did not do are considered as students with better metacognitive skills/regulation (Ozturk, 2017).

vi. Error Detection Paradigm

This method is also used for metacognitive skills. Individuals are presented with texts that contain errors. Students' metacognitive ability is measured from their attention to identify errors in the text. Primary assumption on this is these errors/problems disturb the comprehension of the text and students whose metacognitive monitoring ability is high, will identify these errors while they are reading (Ozturk, 2017).

vii. Measures Used in Computerized Environment

Log file analysis is one method of assessing metacognitive skills in computerized environment. A learning task is performed in the computer and students' activities on the computer is recorded as log data in a separate file. These log file data should be evaluated using a coding system to infer metacognitive activities from these log file data (Veenman et al., 2014).

Eye tracking is also a method that is used to get some impressions about metacognition. No of blinks, eye movements, size of the pupil may reveal some information about metacognition specially metacognitive monitoring (van Gog & Jarodzka, 2013). Relevant information is extracted from the eye movements and transcribe into details about metacognition. Eye tracking data can be coupled with cognitive and metacognitive processes with high sensitivity and specificity and generate algorithms (van Gog & Jarodzka, 2013). They can be adapted in computer environment.

2.3 Online and Offline Methods for Assessing Metacognitive Skills in Mathematical Problem Solving

The previous section examined the available methods for assessing metacognition. It has been recognized that online measures are the optimal choice for evaluating metacognitive skills in Mathematics (Veenman & van Cleef, 2019). Despite online measures yielding reliable data, conducting these methods can pose administrative challenges. Table 2 illustrates a comparison between online and offline measures for assessing metacognition.

Online measures, such as the think-aloud protocol and systematic observations, are time-consuming methods that are not suitable for larger classes. One of the main differences between offline and online measures is that offline measures rely solely on self-reports from individuals, whereas online measures involve coding student behavior based on externally defined criteria (Veenman & van Cleef, 2019).

In the think-aloud method, students verbalize their thoughts while performing the task. Although this may slow down the process, several studies have confirmed that this verbalization does not significantly affect the task performance (Veenman & van Cleef, 2019). In online observations, judges monitor metacognitive behavior either directly by observing the student or indirectly through video recordings. Observations are suitable when the task is not conducive to verbalization or when the student is not proficient enough to describe the task (e.g., primary students) (Veenman & van Cleef, 2019). Similar to the think-aloud method, observations are also time-consuming.

Log file data analysis, performed while the student is engaged in a task on a computer, is another online method. All students' log file data are automatically analyzed to identify the frequency of metacognitive activities and meaningful patterns in the data set (Veenman & van Cleef, 2019).

In offline measures such as self-report questionnaires, students have to answer based on their activities. In offline measures students have to seek help from their memory about their behaviors. They have to recall what they normally do. If the questionnaire is prospective, memory problems increase since they have to recall earlier experiences. This type of measurements can be often disturbed by the memory failures. Offline methods face validity challenges due to the prompting effect of questions, where inquiries may disrupt genuine self-reports and lead to socially desirable responses. Additionally, questions can create an illusion of familiarity with strategies or skills, prompting students to inaccurately label their behavior. This effect is particularly pronounced among students with limited metacognitive knowledge. Another issue arises with questions about activity frequency, requiring comparisons with others and potentially varying individual reference points. While consistent reference points can enhance measurement reliability, differences in these points among students can lead to disparate data (Veenman & van Cleef, 2019). There are some studies which have done some investigations about online and offline measures.

Sarac and Karakelle (2012) studied the relationship between online and offline measures. They measured metacognition using two offline and online measures. Teacher rating scale and a self-report questionnaire were used as offline measures. 47 fifth grade students participated in the study. The online measures were thinking aloud protocol and accuracy ratings of text comprehension. They concluded that using a combination of both online and offline measures will be more appropriate rather than using only online or offline measure.

Online	Offline
1. Conducts while task is doing	1. Conducts after or before the task.
2. Can conduct in smaller classes. Not practical for larger classes.	2. Can conduct in larger classes.
3. Administration is not easy.	3. Administration is easy.
4. Time consuming.	4. Minimum time consumption
5. Data is more reliable.	5. Data distortions can occur due to memory failures.
6. High response rate.	6. Low response rate and poor responses due to comparisons with various reference points such as the top student in class.

Table 2: Comparison on Online and Offline measures

In 2019, Veenman and van Cleef conducted a study with third grade students to identify the best method for measuring metacognitive regulation/skill. There they used 5 instruments. Two prospective questionnaires before the mathematics task (MSLQ-Motivated Strategies for Learning Questionnaire and ILS-Inventory Learning Styles), two online methods (observations and thinking aloud) and one task specific retrospective questionnaire after the mathematics task. Results revealed that prospective questionnaires have poor convergent and predictive validity in measuring metacognitive skills in mathematics. Retrospective questionnaire did slightly better than retrospective questionnaires. Online methods stand out the offline methods. It is concluded that online measures are preferred over offline measures to measure metacognitive skillfulness in mathematics (Veenman & van Cleef, 2019).

Most of the studies confirm that online measures are suitable for measuring metacognitive skills in mathematics. But conducting online measures is creating administrative issues specifically for large classes. It is debatable to decide whether "offline or online". Offline methods such as questionnaires are easy to administer even for large number of students. But online measures, even though they are generating reliable data, are not suitable for large classes because of time and effort waste.

The correlation between self-report questionnaires and online assessments methods such as think aloud method, is reported to be low ranging from -0.07 – 0.31 (Schellings et al., 2013). These low values are reported due to several reasons. Limited self-reporting ability of respondents is the main reasons to have poor responses in self-report questionnaires (Schellings et al., 2013). In addition, different metacognitive activities may be measured by each method. Further, offline measures are not bound to a specific task while online measures are always bound to a specific task such as laboratory experiment, mathematical problem solving etc. (Schellings et al., 2013). When measuring metacognition using online measures, respondents will respond based on the activity they are engaged in. But in offline measures, they have to memorize what they did or what they usually do. That is the task general nature of offline measures such as questionnaires.

2.4 Assessing Metacognition: Questionnaire or Think Aloud?

Questionnaires are commonly used to examine metacognition. Respondents deduce their activities during studying while completing questionnaires. Questionnaires focus on the learner's activities during actual learning. Respondents rate the frequency or importance of different learning activities using Likert-type scales with five or seven points. Many questionnaires target general learning strategies. Questionnaires like the Approaches to Studying Inventory, Study Process Questionnaire, Learning and Study Strategies Inventory (LASSI), and the Motivated Strategies for Learning Questionnaire (MSLQ) reveal how learners usually learn, regardless of the learning environment, task, content, and objectives (Schellings et al., 2013).

Many questionnaires are based on interview-based research, where learners are questioned about conducting learning tasks. Validity of self-reports is assumed to depend on mental episodes persisting in short-term memory for a specific learning task and in long-term memory for a general learning task (Schellings et al., 2013). Questionnaires can vary in specificity, assessing activities in one school domain or specific learning tasks. Mokhtari and Reichard (2002) developed the Metacognitive Awareness of Reading Strategies Inventory (MARSI) for academic reading contexts, assessing students' awareness and perceived use of reading strategies. All questions here ask about “what do you generally do?” in a general reading exercise. Items do not pertain to a specific reading task or a detailed reading assignment (Schellings et al., 2013). There are some efforts on generating task specific questionnaires for different learning tasks. A questionnaire was designed for reading an expository text by Samuelstuen and Bråten in 2007 (2007). Questions are based on the general LASSI instrument. However, these questions were adapted to point to the reading task as a frame of reference (Schellings et al., 2013).

Conducting questionnaires is preferred due to least labor intensive in both administering and processing raw data. Further, questionnaires have reported adequate reliabilities. Questionnaires may vary based on specificity, goals, content and targeted population etc (Schellings et al., 2013). Differences in questions in the questionnaire may lead to low correlations with other measures such as think aloud.

Think aloud is commonly used to assess metacognitive activities in educational research. Participants verbalize their thoughts while they are working on a specific task. These verbalizations are considered as a representation of the underlying thinking process of the current task. These activities are deduced to a well-defined coding system for analyzing. Since individuals have direct access to the working memory, verbalizing their thought process is considered to be more accurate (Schellings et al., 2013). Researchers have found that thinking aloud is a non-reactive measurement method which does not affect the metacognitive activities. It only slows down the process (Schellings et al., 2013). From thinking aloud, unexpected learner activities can be revealed. Further, we can identify new insights from the verbalization of the metacognitive activities. This method reveals information about activities or behaviors that are not yet automated (Schellings et al., 2013). Thinking aloud has a limited scope where non-automated processes in the mind is not revealed. Interpretation of thinking aloud process may require trained raters, which has the possibility of creating interrater issues. Thinking aloud protocol which is used for different tasks may have different evaluation methods and coding systems which may impact the consistency of the method (Schellings et al., 2013).

Even though thinking aloud is recognized as a suitable option for collecting maximum amount of information regarding it may not cover information such as facial expressions and silent moments. When unspoken process is not evaluated, it may create some uncertainty. At the same time, verbal reports may not complete since participants may not reveal their complete thought process.

Hence, both questionnaires and think aloud have their own pros and cons. There are studies which have been conducted to study the relationship between think aloud and questionnaire. In a study by Veenman et al., 2003, questionnaires such as the Inventory Learning Styles (ILS) and the Motivated Strategies for Learning Questionnaire (MSLQ) were administered to students prior to tasks while thinking aloud. Correlations between questionnaire scores and think-aloud measures varied, indicating some inconsistency in the assessment of metacognitive skills (Veenman et al., 2003).

Cromley and Azevedo (2007)conducted a study comparing three parallel strategy-use measures: a general questionnaire (MARSI), think-aloud protocols, and a multiple-choice strategy-use measure. The multiple-choice strategy-use measure required participants to read a passage and choose the best answer from four options, capturing actual strategy use. Both the multiple-choice measure and think-aloud protocols correlated significantly with reading comprehension measures and with each other. However, the MARSI showed non-significant correlations with the other measures, indicating potential limitations in its assessment of strategy use (Cromley & Azevedo, 2007).

Several studies have indicated the superiority of online measures in assessing metacognition, particularly in the context of mathematical problem-solving (Veenman & van Cleef, 2019). Veenman and his team found a moderate correlation between self-report questionnaires and online Think Aloud methods, suggesting the effectiveness of online approaches for evaluating metacognitive skills (Veenman & van Cleef, 2019). Emphasizing the importance of online methods in assessing metacognition, Veenman recommended their preference over offline methods, especially in mathematical contexts (Veenman & van Cleef, 2019). Despite the advantages, conducting the Think Aloud Protocol in larger classes presents several challenges, as highlighted by Schellings and colleagues, who noted its suitability primarily for research purposes rather than practical applications due to its labor-intensive nature (Schellings et al., 2013).

Therefore, employing the Think Aloud method to assess metacognition in Digital Learning Environments (DLEs) proves impractical given the large class sizes. Nonetheless, there remains potential for developing questionnaires based on the Think Aloud method. However, correlations between questionnaire data and Think Aloud measures typically range from moderate to low (Schellings et al., 2013). When comparing questionnaires with Think Aloud, the results vary. Researchers have observed that general questionnaires demonstrate a low correlation (0.22) with Think Aloud, whereas task-specific questionnaires show a moderately high correlation (0.42) (Schellings et al., 2013). Since questionnaires inquire about activities that students performed, they rely on long-term memory. However, they may not accurately represent the actual activities due to memory limitations (Schellings et al., 2013). Responses to questionnaires may be influenced by varying reference points, such as comparing oneself with others, like the teacher or the best/worst student in the class, contributing to the low correspondence of the questionnaire. Even though students may use more strategic activities, they are constrained by the questionnaire provided. Students may report activities as effective not because they use them, but because they believe they are effective. Variations in rating the activities in questionnaires and Think Aloud create non-correspondence between the two. For example, since Think Aloud uses a frequency scale, questionnaires measuring the usefulness of activities do not exhibit a high relationship (Schellings et al., 2013). Due to these reasons, self-report questionnaires and the Think Aloud protocol suffer from low correlations.

Schellings and colleagues (2013) developed a questionnaire to assess metacognition in reading activities, modeled after a taxonomy for coding Thinking Aloud protocols. Although the scale had some validity concerns, it exhibited a promising correlation (r=0.63) with Thinking Aloud. However, there is currently no domain-specific questionnaire for Mathematical problem solving derived from Thinking Aloud. In this study, a questionnaire that was constructed based on the scoring system devised by Veenman and colleagues (2000;2005) for systematic observations utilized in Thinking Aloud during Mathematical problem solving (Athukorala et al., 2024) is compared with Thinking Aloud Protocol to investigate how far this questionnaire is compatible with Thinking Aloud Protocol.

3. Methodology

3.1 Questionnaire Design

As mentioned above, the questionnaire on Measuring Metacognitive skills on Mathematical Problem Solving was developed based on the systematical observations suggested by Veenman and colleagues (Athukorala et al., 2024). It has 12 questions aligned with systematical observations in think aloud protocol. There are three possible answers such as "Often", "Sometimes" and "Almost never" which measures the frequency of each activity in Mathematical Problem Solving similar to Think Aloud evaluation. Student selects the suitable answer from the given 3 answers.

Content validity, construct validity and the reliability were calculated to ensure the results generated by the questionnaire. Content Validity Index (S-CVI/Ave) was 0.9 (Athukorala et al., 2024) and it confirmed the content validity of all the questions. Seven experts from the discipline of Mathematical education answered a questionnaire by studying the relevance, meaning and the correctness of it. Convergent and divergent validity were calculated to confirm the construct validity. Correlation between the questionnaire and the Cognitive and Metacognitive component of MSLQ (Metacognitive Strategies of Learning Questionnaire-35 questions) was calculated and it showed a significant correlation with MSLQ-Cognitive and Metacognitive component (Athukorala et al., 2024). Divergent validity was tested between the questionnaire and the Test Anxiety component of MSLQ. It did not show any significant correlation confirming the divergent validity (Athukorala et al., 2024). Hence, the newly designed questionnaire ensures the validity of the results it generates. Reliability of the questionnaire was calculated using ordinal alpha since the data generated by the questionnaire is ordinal. It was 0.89 endorsing a higher reliability of the results it generates (Athukorala et al., 2024).

As this questionnaire was designed based on the systematical observations on Thinking Aloud protocol, this study expects to study the relationship between the Think Aloud Protocol and the newly designed questionnaire. Methodology followed was described below.

3.2 Resources

Questionnaire on Measuring Metacognitive Skills on Mathematical Problem Solving.

3.3 Data Collection

27 students in third year who are following BSc in Information Technology (Hons) in Sri Lanka Technological Campus, Sri Lanka participated for the study. They first signed a

consent form which indicated their consent on participating the study. Every participant was rewarded for their participation. 20 responses were eligible for the analysis.

3.4 Procedure

Students were given a problem on Linear Congruential Method in Advanced Cryptography module. From the beginning of the solving process to the end, students are advised to verbalize their thought process (Thinking Aloud Activity) and record it. As soon as the solving process is over, students are given the Questionnaire on Measuring Metacognitive Skills on Mathematical Problem Solving.

3.5 Evaluation

Thinking Aloud activity was evaluated by two examiners and the average mark from both evaluators is considered. For that, evaluation criteria in Table 3 was used (Jacobse & Harskamp, 2012). For questionnaire, there were 3 options to be selected which describes the frequency of the activities that students are performing when they solve a Mathematical problem. For the options, “Often, Sometimes and No at all” score of 2, 1 and 0 are given respectively. Then the total marks for both evaluations for each student were calculated. Correlation between the Think Aloud measures and the questionnaire was calculated.

Results are discussed in the next section.

Think Aloud Evaluation			
Activity	**Executed (2 marks)**	**Partially executed (1 mark)**	**Not executed (0)**
(1) entirely reading the problem statement (as incomplete task analysis leads to trail-and-error			
(2) selection of relevant data (task analysis)			
(3) paraphrasing of what was asked for (task analysis and goal setting)			
(4) making a drawing related to the problem (task analysis)			
(5) designing an action plan before actually calculating (planning)			
(6) calculation correctness (avoid sloppiness)			
(7) avoiding negligent mistakes (such as inattentively switching numbers)			
(8) orderly note-taking of problem solving steps (in order to keep an overview of problem-solving steps and create an opportunity for checking outcomes)			
(9) checking the answer			
(10) drawing a conclusion (recapitulating)			

(11) reflecting on the answer (referring to the problem statement)			
(12) relating to earlier problems solved (reflection with the aim to learn from one's experiences).			

Table 3:Marking Rubric for Think Aloud Protocol (Jacobse & Harskamp, 2012)

4. Results

Data from 20 students out of 27 (74%) were able to use for results analysis. Each student's score for both questionnaire (Q) and think aloud (TAP) activity was used for the analysis. Using Statistical Package for Social Science (SPSS) software, correlation between each component of metacognitive skills (planning, monitoring and evaluation) and overall correlation was calculated. Data generated from both methods are ordinal. Hence, correlation was calculated using Spearman's rho. The Table 4, Table 5 and table 6 show the correlation coefficient of each component.

Correlations				
			PlanningTAP	PlanningQ
Spearman's rho	PlanningTAP	Correlation Coefficient	1.000	.530*
		Sig. (2-tailed)	.	.016
		N	20	20
	PlanningQ	Correlation Coefficient	.530*	1.000
		Sig. (2-tailed)	.016	.
		N	20	20
*. Correlation is significant at the 0.05 level (2-tailed).				

Table 4: Correlation of Planning Component

Correlations				
			MonitoringTAP	MonitoringQ
Spearman's rho	MonitoringTAP	Correlation Coefficient	1.000	.530*
		Sig. (2-tailed)	.	.016
		N	20	20
	MonitoringQ	Correlation Coefficient	.530*	1.000
		Sig. (2-tailed)	.016	.
		N	20	20
*. Correlation is significant at the 0.05 level (2-tailed).				

Table 5: Correlation of Monitoring Component

Correlations				
			EvaluationTAP	EvaluationQ
Spearman's rho	EvaluationTAP	Correlation Coefficient	1.000	.500*
		Sig. (2-tailed)	.	.025
		N	20	20
	EvaluationQ	Correlation Coefficient	.500*	1.000
		Sig. (2-tailed)	.025	.
		N	20	20
*. Correlation is significant at the 0.05 level (2-tailed).				

Table 6: Correlation of Evaluation Component

The Table 7 shows the overall correlation between the questionnaire and the think aloud protocol. It is 0.517 with 0.05 significant level.

Correlations				
			TAP	Q
Spearman's rho	TAP	Correlation Coefficient	1.000	.517*
		Sig. (2-tailed)	.	.020
		N	20	20
	Q	Correlation Coefficient	.517*	1.000
		Sig. (2-tailed)	.020	.
		N	20	20

*. Correlation is significant at the 0.05 level (2-tailed).

Table 7: Overall Correlation between the Questionnaire and Think Aloud

5. Challenges and Limitations

It was challenging to find students willing to participate in the think-aloud exercises. Because they had to verbalize their thought processes while solving problems, many students hesitated and showed reluctance to take part. Eventually, 27 students agreed to participate, but only 20 recordings met the required conditions. Issues such as incomplete recordings, recordings that did not sufficiently describe the entire thought process, and unclear recordings contributed to the low number of usable responses for the study.

6. Discussion

Metacognitive skills play a significant role in mathematics (Özcan, 2014). Assessing the metacognitive skills of each student in the classroom can provide valuable opportunities for both students and teachers to enhance the teaching-learning process. Veenman and colleagues (2019) conducted a study comparing online and offline measures of metacognitive skills, concluding that online measures are preferable to offline measures (Veenman & van Cleef, 2019). However, implementing online assessment methods in the classroom poses challenges due to time and effort constraints.

Offline measures are easier to administer, but their reliability in generating accurate information is a concern. Therefore, it is essential to investigate the relationship between offline measures and online measures. This study aimed to examine the correlation between an offline measure, the Questionnaire on Measuring Metacognitive Skills on Mathematical Problem Solving (Q), and the Think-Aloud Protocol (TAP). The results revealed a moderate correlation of 0.52 between the questionnaire and the think-aloud protocol, indicating a moderate level of agreement between the two measures.

Furthermore, the study compared three main components—planning, monitoring, and evaluation—between the questionnaire (Q) and the think-aloud protocol (TAP). All three components showed moderate correlations with the think-aloud method.

7. Conclusion

Due to practical challenges in implementing online assessment methods to measure metacognitive skills in classrooms, there is a pressing need to develop offline measures that can generate reliable information comparable to online measures. Online methods like the think-aloud protocol, systematic observations, etc., are suitable for research purposes and small classes. However, they are not feasible for large classes or massive online courses. In such scenarios, offline measures that have been validated against online measures and demonstrate comparable reliability play a crucial role. The overall correlation between the questionnaire and TAP (r=0.52) suggests the potential utility of this questionnaire as an alternative to TAP.

References

Akturk, A. O., & Sahin, I. (2011). Literature Review on Metacognition and its Measurement. *Procedia - Social and Behavioral Sciences*, *15*, 3731–3736. https://doi.org/10.1016/j.sbspro.2011.04.364

Athukorala, U., Fernando, D., & Wijeratne, C. (2024, February 13). Thinking Aloud Protocol Based Self-Report Questionnaire to Measure Metacognitive Skills in Mathematical Problem Solving. *The Asian Conference on Education 2023 Official Conference Proceedings*. The 15th Asian Conference on Education, Japan.

Cromley, J. G., & Azevedo, R. (2007). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning*, *1*(3), 229–247. https://doi.org/10.1007/s11409-006-9002-5

Dabbagh, N., & Kitsantas, A. (2005). Using Web-based Pedagogical Tools as Scaffolds for Self-Regulated Learning. *Instructional Science*, *33*(5–6), 513–540. https://doi.org/10.1007/s11251-005-1278-3

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

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, *94*(2), 249–259. https://doi.org/10.1037/0022-0663.94.2.249

Ozturk, N. (2017). Assessing Metacognition: Theory and Practices. *International Journal of Assessment Tools in Education*, 134–134. https://doi.org/10.21449/ijate.298299

Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *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*. 15.

Schellings, G. L. M., Veenman, M. V. J., Van Hout-Wolters, G., & 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*. https://doi.org/10.1007/s10212-012-0149-y

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting Self-Regulation in Science Education: Metacognition as Part of a Broader Perspective on Learning. *Research in Science Education*, *36*(1–2), 111–139. https://doi.org/10.1007/s11165-005-3917-8

van Gog, T., & Jarodzka, H. (2013). Eye Tracking as a Tool to Study and Enhance Cognitive and Metacognitive Processes in Computer-Based Learning Environments. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (Vol. 28, pp. 143–156). Springer New York. https://doi.org/10.1007/978-1-4419-5546-3_10

Veenman, M. V. J., Bavelaar, L., De Wolf, L., & Van Haaren, M. G. P. (2014). The on-line assessment of metacognitive skills in a computerized learning environment. *Learning and Individual Differences*, *29*, 123–130. https://doi.org/10.1016/j.lindif.2013.01.003

Veenman, M. V. J., Prins, F. J., & Verheij, J. (2003). Learning styles: Self-reports versus thinking-aloud measures. *British Journal of Educational Psychology*, *73*(3), 357–372. https://doi.org/10.1348/000709903322275885

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

Contact email: uthpalap@itum.mrt.ac.lk