

On-line and Off-line Assessment of Metacognition

Seda SARAÇ*

Yildiz Technical University, Turkey

Sema KARAKELLE

Istanbul University, Turkey

Received: December 2011 / Revised: February 2012 / Accepted: March 2012

Abstract

The study investigates the interrelationships between different on-line and off-line measures for assessing metacognition. The participants were 47 fifth grade elementary students. Metacognition was assessed through two off-line and two on-line measures. The off-line measures consisted of a teacher rating scale and a self-report questionnaire. The on-line measures were thinking aloud protocols and accuracy ratings of text comprehension. The results showed positive significant correlation between data from two off-line measures and negative significant correlation between data from two on-line measures. The off-line metacognitive measures had non-significant correlations with all on-line measures. Principal Component Analysis, performed on four metacognitive measures, yielded a two-factor solution and this two-factor solution accounted for 71.5 % of the sample variance. The data from two off-line measures loaded on the first component with a variance proportion of 38.6 % and the data from two on-line measures loaded on the second component with a variance proportion of 32.9%. The findings of the study showed that metacognitive processes form a complex structure that needs to be assessed using various methods. However, in the multi-method studies, using on-line and off-line measures together will be appropriate rather than using only on-line measures or only off-line measures.

Keywords: Metacognition, on-line/off-line assessment, think aloud, accuracy rating, self-report, teacher ratings.

Introduction

Over the past thirty years, there has been growing interest among researchers in the study of metacognition. Flavell (1979), defined metacognition as "the individuals' knowledge about cognitive processes and the application of this knowledge for controlling the cognitive process". Metacognition has been postulated as a multifaceted and overarching structure

*  Seda Sarac, Yildiz Technical University, Faculty of Education, Istanbul, Turkey. Phone: +90 535-4144175 Fax: +90 212-3834808. E-mail: sbiryan@yahoo.com

made up of sub-elements, each having different features. Flavell (1979) classified metacognition into two dimensions, as metacognitive knowledge and metacognitive experiences; and Brown subsumed metacognition under two dimensions namely, knowledge of cognition and regulation of cognition. In both classifications, the second dimensions are defined similarly as an individual's monitoring, controlling and regulating his/her own cognitions. According to Efklides (2006; 2009) metacognition has three subcomponents, namely, metacognitive knowledge, metacognitive experiences and metacognitive skills. Recent research shows that metacognition is taken as a three-faceted structure being *metacognitive knowledge*, *metacognitive monitoring* and *metacognitive control* (Dunlosky & Metcalfe, 2009).

Metacognitive knowledge involves knowing one's own cognitive characteristics (knowledge of person), the nature of different cognitive tasks (knowledge of task) and the possible strategies that enable the fulfilment of different cognitive tasks (knowledge of strategy). Because metacognitive knowledge is stored in the long-term memory, by nature, it is relatively static and declarative knowledge (Flavell, 1979; 2000). *Metacognitive monitoring* refers to assessing or evaluating the ongoing progress or current state of particular cognitive activity. *Metacognitive control* pertains to regulating on ongoing cognitive activity (Dunlosky & Metcalfe, 2009).

It's obvious that how metacognition is modelled is closely related to both the assessment methods of metacognition and the results that are concluded from these assessments about the metacognitive processes. For this reason, it is important to examine the methods for assessing metacognition. Metacognition can be assessed by many different methods. These methods are usually classified as on-line or off-line according to when they are collected (Desoete, Roeyers & De Clercq, 2003; Pintrich, Wolters & Baxter, 2000; Veenman, 2005).

On-line Measures

On-line measurements are collected while the individual is engaging a specific task in hand. They assess domain specific metacognition with a focus on the learning process. Typically, individuals are recorded during task performance. Think-aloud protocols, accuracy ratings and systematic observations are on-line measures frequently used to assess metacognition.

In *think-aloud protocols*, individuals are instructed to think aloud while they are working on a specific cognitive task. The researcher interferes as little as possible. All utterances are recorded on audio or video-tape. Afterwards, the recordings are transcribed and metacognitive activities are scored according to a coding scheme (e.g. Cromley & Azevedo, 2006; Pressley & Afflerbach, 1995; Thomas & Barksdale- Ladd, 2000; Veenman & Beishuizen, 2004; Veenman, Elshout & Meijer, 1997; Veenman, Kok & Blöte, 2005; Veenman & Veheij, 2003). *Accuracy ratings* refer to ongoing assessments of learning or performance. In this methodology, the individual performs a criterion task and immediately makes a judgement regarding confidence, ease of solution judgements or performance accuracy (Schraw, 2009). The absolute difference between an individual's rating and her/his actual performance is calculated (e.g., Hacker, Bol & Bahbahani, 2008; Hacker, Bol, Horgan & Rakow, 2000; Nietfeld, Cao, & Osborne, 2005; Pressley & Ghatala, 1989). In *systematical observation*, data is collected during individuals' task performance. The judges observe the individual during task performance and/or watch videotapes afterwards and score the individual's metacognitive behaviours (e.g., Veenman, Kerseboom & Imthorn, 2000; Veenman, Kok & Blöte, 2005).

Off-line Measures

Unlike on-line measures, off-line measures aim at assessing metacognition either in general (i.e. without any explicit reference to a specific task) or specific to a task. Task-specific off-line

measurements are collected retrospective to task performance. Common off-line techniques are self-report questionnaires, interviews, and teacher ratings.

Self-report questionnaires are usually Likert type scales, developed with the aim of assessing metacognition. Generally, two types of metacognitive questionnaires are used in metacognition research: general and domain specific. General metacognitive questionnaires are designed to assess metacognition independent of any specific domain (e.g., Pintrich, Smith, Garcia, & McKeachie, 1991; Schraw & Dennison, 1994; Sperling, Howard, Miller & Murphy, 2002). Domain specific self-report questionnaires are generally developed with the aim of assessing metacognition in a single domain such as reading, problem solving, etc. (e.g., Mokhtari & Reichard, 2002; Schmitt, 1990; Fortunato, Hecht, Tittle, & Alvarez, 1991). Another off-line technique to assess metacognition is *interview protocols*. Mainly, there are three varieties of interview protocols encountered in metacognitive research. One way of assessing metacognition using the interview protocols is to simply ask the subjects to describe what is typical behaviour under certain circumstances (e.g., Myers & Paris, 1978; Paris & Jacobs, 1984). Alternatively, individuals are asked to describe their metacognitive behaviours after completing a specific task (e.g., Artzt & Armour-Thomas, 1992). In more advanced interview protocols, hypothetical learning situations are depicted and subjects are asked what they would do in these particular situations or they are asked to generate as many possible strategies that can be used in such situations as they can think of (e.g., Annevirta, Laakkonen, Kinnunen & Vauras, 2007; Zimmerman & Martinez-Pons, 1988, 1990). *Teacher ratings* are another off-line way of assessing metacognitive levels of school-age children. The teachers are requested to evaluate their students' metacognition on a rating scale (e.g., Bingham & Whitebread, 2008; Desoete, 2008; Sperling, Howard, Miller & Murphy, 2002; Whitebread & Coltman, 2010; Whitebread et al., 2009).

Although studies in this area are increasing exponentially, it is observed that there are still issues related to measuring metacognition (Winne & Perry, 2000; Veenman, 2005). These issues are not only limited to the development of various techniques aimed at measuring metacognition but also the need to analyse the correlation between these techniques as well as their validity, reliability (Schraw, 2009).

Relations among Metacognitive Measures

Results from several studies using multiple metacognitive measures discredit the measures that are frequently used in metacognitive research and compel researchers to scrutinize what they are actually measuring. For instance, in studies using multiple on-line measures, in general, significant correlations are reported between measurement methods. Veenman, Kerseboom and Imthorn (2005) and Veenman, Kok and Blöte (2005) examined the metacognitive skills of 12- and 13-year-olds using on-line systematic observation and think-aloud protocol analyses. They reported a significant correlation between the assessment methods ($r = .78$ and $r = .89$, respectively). Veenman, Wilhelm and Beishuizen (2004) reported that the results of the logfile and think-aloud protocols in university students yielded high correlation values. Log file measures and protocol analysis correlated .85 for the task in the domain of biology and .84 for the task in the domain of geography with one another. Cromley and Azevedo (2006), in their study of ninth-grade students, found significant correlations between the scores from think-aloud protocol analysis and the scores from concurrent multiple-choice strategy use measure.

On the other hand, mixed results are obtained from studies using multiple off-line measures. Minnaert and Janssen (1997), in their study with college students, compared the results from two questionnaires: the Leuven Executive Regulation Questionnaire and the Inventory of Learning Styles. The researchers found correlations between .13 to .80 between

corresponding subscales of the questionnaires. In their study, Sperling, Howard, Miller and Murphy (2002) examined the correlations among various questionnaires (Jr. Metacognitive Awareness Inventory, Index of Reading Awareness, Metacomprehension Strategies Index and Strategic Problem Solving) for assessing metacognition of students in 3rd through 9th grades. They found no substantial correlations among the results from the questionnaires. The researchers also compared the questionnaires' results to teacher ratings. They found significant correlation for the younger group but not the older group. In the same vein, Sperling, Howard, Stanley and DuBois (2004) investigated college students' metacognition in two studies. In their first study, they assessed college students' metacognition using two questionnaires, namely, Metacognitive Awareness Inventory and Learning Strategies Survey. Results from the two questionnaires correlated .50 with one another. In their second study, the researchers examined the correlations between results from Metacognitive Awareness Inventory and Motivated Strategies for Learning Questionnaire. The two questionnaires correlated significantly with one another. However, in a study with 3rd graders, Desoete (2008) reported that there was no significant correlation among scores from the prospective questionnaire, the retrospective questionnaire and the teacher ratings.

The results of the studies employing off-line and on-line techniques in combination generally show that there is no significant correlation among scores from off-line measures and on-line measures. In their studies with college students, Schraw and Dennison (1994) and Sperling, Howard, Stanley and DuBois (2004) found no substantial correlation between college students' monitoring accuracy scores and the results from the Metacognitive Awareness Inventory. However, Schraw (1997) reported a significant correlation of 0.30 between monitoring accuracy and monitoring strategies in college students. Studies concerning young age groups have also revealed similar results. Hannah and Shore (1995) analysed the metacognitive skills of primary and secondary school students using a think-aloud protocol and prospective interviewing and reported that there was non-significant correlation between the two measures ($r = .26$). In a study on 3rd and 4th graders, Van Kraayenoord and Schneider (1999) reported that there was non-significant correlation between the results of the qualitative protocol analyses and the results of the reading strategies questionnaire in both grades ($r = .26$ for third graders and $r = .07$ for fourth graders). In their study with ninth grade students, Cromley and Azevedo (2006) reported that the scores from the self-report questionnaire did not correspond neither with the scores from think-aloud protocol analysis nor with the scores from concurrent multiple-choice strategy use measure. In Desoete's (2008) study with third graders, too, the scores from two off-line questionnaires did not correlate with the scores from think-aloud protocols.

When the results obtained in the studies mentioned above are considered together, we see that the results obtained by means of different on-line methods are related with each other. Likewise, there are relations among the results obtained from off-line methods. However, there is not a relationship between the scores obtained by means of off-line methods and on-line methods. In his comprehensive review, Veenman (2005) also showed that scores from off-line measures do not correspond to individuals' scores from actual behavioural measures during task performance. In other words, data from the off-line and the on-line measures, generally, do not correlate with each other.

However, the majority of studies, whose results have been addressed together, have used various types of off-line and on-line techniques and various types of criterion tasks. This variety confounds the conclusion of precise results about the validity and reliability of the measurements. We believe that examining the participants' metacognition within the same criterion task and by using more than one on-line and one off-line measure will better reveal the relations between the measures. Along these lines, in this study our aim is to compare

two on-line and two off-line methods in relation to a text-learning task. At the same time, the study aims to identify the patterns between the measures by conducting a factor analysis.

Methods

Participants

The participants were from three state schools in Istanbul. The schools were purposefully selected because they educate children mostly from families with average income, judged by the opinions of the school principals and classroom teachers. The students were selected randomly from six classes (two classes from each school and 10 students from each class). The total number of the participants was 60. Think aloud protocols of 4 students could not be transcribed due the high noise level in the background, as some parts of the protocols coincided with the break time. From 56 students, all the students with missing data were eliminated. Eventually, the participants in this study were 47 fifth graders (20 girls, 27 boys, $M_{age} = 10.00$ years, age range: 9-11 years).

The teachers who participated in the study were the classroom teachers of these six classes. In the first five years of compulsory education in Turkey, students remain with the same teacher. In some rare cases, the teacher can be changed due to illness, school change, retirement, etc. However, the participant teachers in this study had been teaching the same students for five years. The average professional experience of the participating teachers was 11.5 years.

Measures

Off-line measures. Two off-line measures of metacognition were used in this study.

Jr MAI (Form A). The Turkish version of the Jr. Metacognitive Awareness Inventory- Form A was used for the study (Sperling et al., 2002). Jr. Metacognitive Awareness Inventory-Form A (Jr. MAI-A), a self-report inventory, was developed as a measure of general metacognitive awareness of children in grades 3-5. The Jr. MAI was developed from a previous instrument, the Metacognitive Awareness Inventory (MAI), used with adult populations (Schraw & Dennison, 1994). Jr. MAI is a 3-point likert type scale ranging from 1 ("never") to 3 ("always"). Its purpose is to assess children's domain general metacognition. The original inventory consists of 12 items ($\alpha=.76$) with two subscales, namely, the knowledge of cognition (e.g., "I learn more when I am interested in the topic") and the regulation of cognition (e.g., "When I am done with my schoolwork, I ask myself if I learned what I wanted to learn"). Although originally there were two subscales as the results of the factor analysis yielded a single factor solution, the researchers recommended using the inventory as an overall measure of metacognition.

The Turkish version of Jr. MAI was adapted by Karakelle and Saraç (2007). The Turkish version of the inventory consisted of 12 items. The internal consistency reliability for the scale was .64 and test-retest reliability of the Turkish inventory was .74 ($N = 356, p < .01$). The factor analysis for the Turkish version yielded one factor solution; the authors recommended using the scale as an overall measure of metacognitive awareness. For this study, the internal consistency reliability of the scale was .70. Jr. MAI is chosen since it is the only metacognition scale adapted for Turkish samples in this age group.

Teacher rating scale. A rating scale that is adapted from Sperling et al. (2002) was used to collect teachers' opinions about the students' metacognition. Prior to rating, the teachers were provided two information sheets, one with a brief explanation of metacognition and typical characteristics of metacognitive children and the other with behavioural descriptors to distinguish students who are high in metacognition (e.g. "judges performance accurately",

“asks questions to insure understanding while learning”). After reading the information sheets mentioned above, the teachers then rated each of their students accordingly, on a scale ranging from 1, designating “very low metacognition” to 6, designating “very high metacognition”. No significant differences by teacher were indicated in the ratings, $F(5, 41) = 0.351, p < .001$.

On-line measures. Two on-line measures of metacognition were used for the study.

Think-aloud protocols. Students were presented a text-learning task. The text for this study, taken from Demirel (1995), was about the design, working principles and types of balloons. The text consisted of nine paragraphs with 456 words. Prior to the study, seven fifth grade teachers, other than the participating teachers, read the text and judged it as appropriate for fifth grade readers.

The children were instructed to think aloud while studying the text. All the readers' utterances were audiotaped and transcribed. All the transcriptions were segmented according to a study by Cote, Goldman & Saul (1998) in which unit of analysis was defined as “a comment or set of comments on the same core sentence or group of sentences as well as the reading behaviour associated with those comments” (p. 14). After the identification of the units of analysis, the units were analyzed according to the Taxonomy of Metacognitive Activities in Text-studying (TMATS), developed by Meijer, Veenman and van Hout-Wolters (2006). TMATS consists of five categories: orientating, planning, executing, evaluating and elaborating. Under each category, there are several metacognitive activities. The total number of metacognitive activities listed in the taxonomy is 70. After several analyses, Meijer, Veenman and van Hout-Wolters (2006) concluded that the more parsimonious distinction of Flavell (1979) would be more suitable, so the researchers reverted to the original three categories. The activities of orientating and planning were subsumed under the category of planning, the activities of monitoring were subsumed under the category of monitoring, and the activities of evaluating and elaborating were subsumed under the category of evaluating. The category of executing was left out as most activities in this category were thought to reflect cognitive activities rather than metacognitive activities. According to the taxonomy, the category of planning, combined with orientating, consisted of 15 metacognitive activities (e.g. establishing task demands, continue reading hoping for clarity, selecting a particular section of text to look for required information). The category of monitoring consisted of 12 metacognitive activities (e.g. noticing unfamiliar terms or words, commenting on task demands, noting lack of knowledge). The category of evaluating, combined with elaborating, consisted of 12 activities (e.g. finding similarities, explaining strategy, connecting parts of text by reasoning). For the entire taxonomy of metacognitive activities in text studying, see Meijer, Veenman and van Hout-Wolters, 2006.

Three judges, knowledgeable in metacognition and reading processes, segmented the protocols simultaneously. The three judges scored all the protocols, independently, on the presence of metacognitive activities in TMATS. Each unit, corresponding to a metacognitive activity on the taxonomy, was coded in the margin as belonging to one of the three categories: planning (e.g. “I’m going to read this part about valves again”), monitoring (e.g. “I don’t know what this word means”) and evaluating (e.g. “I’m glad that I read this part again because now I understand what it says”). Then, for each student, the number of activities under each category was counted. Three scores (planning, monitoring and evaluating) were computed for each student. Table 1 shows the descriptive statistics for the categories of TMATS. The interrater reliability was 96% between the first and the second judge, 97% between the first and the third judge and 96% between the second and the third judge.

Table 1. Descriptive statistics for the categories of metacognitive activities

<i>N</i> = 47	<i>M</i>	<i>SD</i>	Minimum	Maximum
Planning	1.45	1.87	0	8
Monitoring	1.34	2.71	0	16
Evaluating	7.61	8.00	0	32

Accuracy Ratings. Accuracy measures the degree to which children's confidence judgments match their actual test performance (Hacker, Bol & Bahbahani, 2009; Hacker, Bol, Horgan & Rakow, 2000; Pressley & Ghatala, 1989). Metacognitive monitoring accuracy was calculated by taking the absolute value of the difference between students' ratings on the prediction scale and their performance. In this study, the students' performance was assessed by a post-test consisting of 15 multiple choice questions ($\alpha = .77$). Six of the questions were text-implicit and 9 of the questions were text-explicit. The students' prediction judgements (JOL) were used to measure metacognitive monitoring accuracy. After the children studied the experimental text, they were asked to rate how well they think they understood the text on a rating scale ranged from 1, designating "not at all", to 4, designating "very well". For each reader, the difference between rating on the prediction scale (converted into percentages) and performance score (converted into percentages) was calculated and the absolute value of this difference was taken. With this formula, the accuracy scores ranged between 0 and 100, with the scores of 0 indicating perfect accuracy and scores of 100 indicating total inaccuracy. To prevent any confusion due to reverse points, all scores were subtracted from 100 and consequently the accuracy scores for this study ranged between 0 and 100, with the scores of 100 indicating perfect accuracy and scores of 0 indicating total inaccuracy.

Procedures

The first author, in a quiet room in the school, assessed all students individually during school time. In a typical session, at the very beginning, the researcher had a short chat with the student, trying to make the student feel comfortable and safe with the researcher. After this socializing, the child, following the suggestions of Ericsson and Simon (1993), was first instructed to think aloud while working on a text. In this instruction session, two texts, other than the experimental text, were used. The trial session with the trial texts lasted till the subject felt comfortable with thinking aloud (approximately 10-min.). Then the experimental text was introduced. The students were allowed to study the text without any time limit. The shortest think aloud session took 182 seconds period and the longest session took 1494 seconds ($M = 594.09$ $SD = 275.65$). The experimenter used only standard prompts, "Please, keep on thinking aloud" and "What are you thinking?" whenever the student fell silent. No other interaction between the student and the experimenter was allowed. After the students mentioned that they were ready for the test, they were instructed to rate their understanding on the rating scale below the text. Then they were presented with the learning performance test. At the end of the session, the students completed Jr. MAI (Form A). In each school the first author, after finishing data collection with the students, met the classroom teachers individually in a quiet room. After a short introduction about the aims of the study, the teacher was presented with the information sheet about metacognition and requested to read it. The teacher was allowed to read the sheet without any time limit and to ask any questions regarding metacognition. After the reading session, the teacher was asked to rate the participating students from his/her class accordingly on the rating scale. The teacher was requested to base her/his judgements according to the students' typical learning behaviours across domains.

Results

In this study, two off-line and two on-line metacognitive measures were used. Descriptive statistics for each metacognitive measure are presented in Table 2.

Table 2. Descriptive Statistics for Variables

N= 47	M	SD	Minimum	Maximum
Jr-MAI	31.55	2.71	25.00	36.00
Teacher-Rating	4.21	1.35	2.00	6.00
Think-Aloud Protocols	17.02	10.93	1.00	43.00
Accuracy Ratings	78.01	14.63	40.00	100.00

Pearson product-moment correlation coefficients were computed to investigate the interrelations among metacognitive measures. The two off-line measures, Jr. MAI (Form A) and the teacher ratings, correlated significantly with one another ($r = .50, p < .01$). The two on-line measures, think aloud protocols and monitoring accuracy, correlated significantly with one another but the correlation was negative ($r = -.30, p < .05$). No significant correlation was found between the results from Jr. MAI (Form A) and the results from two on-line measures. Also, no significant correlation was found between the teacher ratings and the results from the two on-line measures. Correlations among metacognitive measures are presented in Table 3.

Table 3. Correlations between Jr MAI-A; Teacher Ratings, Think-aloud Protocols and Accuracy Ratings

N= 47	Jr MAI	TR	TAP	MA
Jr MAI-A	1			
Teacher Rating(TR)	.50**	1		
Think-aloud Protocols (TAP)	.12	.12	1	
Monitoring Accuracy(MA)	.07	.21	-.30*	1

** $p < .01$, * $p < .05$

A principal component analysis (PCA) was performed on the results from four metacognitive measures to investigate the factor structure. Previous research provides a wide range of recommendations regarding the sample size in PCA. As a general rule of the thumb at least 300 cases are required for PCA (Tabachnick & Fidell, 1996). However, research has demonstrated that this general rule for minimum sample size is not valid. Sapnas and Zeller (2002) report that the sample size should not be too large and sometimes additional subjects waste research resources. According to MacCallum et al. (1999), the sample size is dependent on the characteristics of the variables and the study. Particularly, the level of variable communalities is important in establishing sample size. High variable communalities, that is, .60 and greater, require small sample size. In the same vein, Wieringa (2009) recommends that in case of high factor loadings and low number of factors, a sample size below 50 is sufficient for PCA.

In this study, there are only two factors and the item communalities range between .63 and .80. So, this sample size of 47 seems to be sufficient for PCA. Furthermore, KMO coefficient (.44) and Bartlett test of sphericity (21.567, $p < .001$) were performed and the results showed that the data is suitable for PCA.

The PCA analysis yielded a two-factor solution. The two factors, with eigenvalues above 1, together accounted for 71.5 % of the total variance explained. The unrotated solution showed that the scores from the two off-line measures loaded on the first factor. This factor, with an eigenvalue of 1.54, accounted for the 38.6 % of the total variance explained by this solution. The scores from the two on-line measures loaded on the second factor. This factor, with an eigenvalue of 1.31, accounted for the 32.9% of total variance explained by this solution. Factor loadings from unrotated solution are presented in Table 4.

Table 4. *Unrotated Component Matrix for Metacognitive Measures*

	Component 1	Component 2
Eigenvalue	1.54	1.31
Teacher Ratings	.85	.27
Jr. MAI	.84	-.01
Accuracy Ratings	.03	.86
Think-aloud Protocols	.22	-.71

Discussion

This study examined the patterns of the relations between metacognition scores obtained via two on-line and two off-line measurement methods.

Relation between Off-line Methods

In the study, a self-report measure (Jr-MAI) and the teacher ratings were used as off-line measures. The results revealed that these two measures are significantly correlated; in other words, the individual's assessment of his own metacognitive activities is compatible with his teacher's assessments which are built on the teacher's observations across domains. Similarly, in the study conducted by Sperling, Howard, Miller and Murphy (2002), a significant correlation between Jr. MAI and the teacher ratings was observed for 3rd, 4th and 5th graders. However, the researchers reported non-significant correlation between Jr. MAI and the teacher ratings for the older age group (6th to 9th graders). Desoete (2008), too, reported a significant correlation between the teacher ratings and the other metacognitive measures for 3rd graders, indicating that the teacher ratings could be an alternative method for metacognitive macro-evaluation. The results gathered from these studies can be interpreted as showing that teacher ratings are more accurate for the young age groups. The fact that the observations of the teachers were consistent with the task-specific observations of the researchers in a study carried out by Whitebread et al. on the preschool children points that the teachers' ratings are appropriate for the young age groups (Bingham and Whitebread, 2008; Whitebread et al., 2007; Whitebread et al., 2009; Whitebread and Coltman, 2010). From this point of view, by carrying out developmental studies in which the teacher observations will be used, it will be possible to explain why the teachers make more accurate evaluations for the younger age groups. In these kinds of studies, it could be interesting to analyse the type of observations that teachers make to assess their students' metacognitive levels. For instance, if the teacher ratings are based on all of the procedural behaviours observable in daily activities, this could bring a whole new perspective to the analysis of metacognitive processes.

The significant correlation between the questionnaire and the teacher ratings also suggests the need to address the criticism that the self-report questionnaires consist of the individual's opinions about one's self. Even if the questionnaires assess the individual's opinions of her own metacognitive activities, in this example, it can be considered that these

opinions do not solely consist of the individual's assumptions as the assessment is supported by an external measure (the teacher ratings).

Relation between On-line Methods

In this study, a think-aloud protocol and accuracy ratings were used as on-line measures for the text-learning task. The results show that the two on-line measures have a significant negative correlation, that is to say, those who can make accurate judgments of learning (JOL) perform less metacognitive activity.

One explanation may be that this negative correlation is the result of underconfidence-with-practice-effect (UWP), a phenomenon introduced by Koriat, Sheffer and Ma'ayan (2002). This effect points out that the JOL accuracy decreases as the amount of practice increases. There are studies showing that the UWP effect is present for both the item by item JOL accuracy and global judgements (Finn & Metcalfe, 2007; Koriat, Ma'ayan, Sheffer & Bjork, 2006, Rast & Zimprich, 2009; Serra & Dunlosky, 2005). In the think-aloud protocols, since the participant continues the learning activity until mastery, he/she generally makes repetitions more than once and thus has the probability to carry out more metacognitive activities. If the JOL accuracy shows a decrease depending on the repetitions, it is a logical result that monitoring accuracy decreases as the metacognitive activity number determined with think-aloud increases.

An alternative explanation of this result may be in terms of study-time allocation. Given that the task used in this study was a text-learning task, lower metacognitive activity means that a short amount of time was allocated to studying the text. Within this framework, those that make accurate judgements of learning could be performing less metacognitive activity because they can correctly eliminate easy-to-learn from difficult-to-learn, and well-learned from to-be learned. This elimination could help the learner use his time more effectively, thus avoiding unnecessary and ineffective strategies. According to Metcalfe (2009) recent studies indicate a causal relationship between JOLs and study behaviour and a negative correlation between time allocated for studying and JOLs. Although this finding is obtained from the studies in which the JOL accuracy is examined item by item, we can expect to obtain similar results for the global judgements. In this direction, it will be appropriate to examine the participants' global judgements in the studies to be carried out relevant to the study-time allocation.

Interrelations among Off-line and On-line Methods

This study did not reveal a significant relationship among the on-line and off-line methods used. These findings are compatible with several of the aforementioned studies. The results of the exploratory factor analysis showed that the metacognitive measures used in the study clearly fall into two distinct categories, namely on-line and off-line methods. The off-line measures are grouped in one single factor, explaining the 38.6% variation in the metacognitive scores. Similarly, the on-line measures are grouped in one single factor, explaining the 32.9% variation in the metacognitive scores. These results suggest that off-line and on-line measures form distinctive assessment structures and these assessment structures are internally coherent. This result can be a sign that off-line and on-line measures assess independent structures that are internally coherent. When developing any type of test, in such a case, each factor would be named separately given that each factor assesses an independent dimension. However, since the study only analyses different methods that aim to assess the same structure, this raises the question of how to explain the measures acting as if they belong to different dimensions. Of course it is possible to explain this discrepancy between assessments using off-line and on-line methods as the weaknesses of the measures. However, this differentiation can also be explained in terms of the elements of

metacognition. The compatibility or incompatibility between studies using multi-methods can be attributed to issues related to the assessment of different structures. As stated by Brown (1987) there are activities that individuals do not need to express verbally or be aware of during the execution of performance. Metacognitive judgements have implicit and unconscious aspects just as they have explicit and conscious ones. According to Koriat (2000, 2008), while knowledge-based metacognitive judgements are built on explicit and conscious inferential processes, experience-based judgements are built on subjective feelings, like the “tip-of-the-tongue” phenomenon. Within this framework, it can be considered that off-line methods such as questionnaires could be more sensitive to explicit and conscious processes and on-line methods such as think- aloud protocols could be more sensitive to implicit and unconscious processes. As also addressed by van Hout-Wolters (2000, cited in Helms-Lorenz & Jacobse, 2008), off-line measures may not clearly reflect the learning activities and off-line and on-line measures tend to measure somewhat different constructs.

In summary, the findings of this study support the view that metacognitive processes form a complex structure that needs to be assessed using various methods. However, in the multi-method studies, using on-line and off-line measures together will be appropriate rather than using only on-line measures or only off-line measures. Using these two methods together in a complementary way may allow us to see the metacognitive activities as a whole. Although there is criticism of the off-line methods, according to the study results, the fact that there is compliance between the teachers’ opinions and the opinions their students hold about themselves reminds us that more consideration is necessary before putting away the questionnaires while deeming them “quick and dirty”. In addition, it should be taken into consideration that the teacher ratings can be an assessment method to be used especially for studying young age groups.

• • •

Seda Sarac is currently in private practice teaching self-regulated learning strategies to children who have learning difficulties. She worked as a research assistant in the Faculty of Education at Yildiz Technical University. Her research interests are metacognition, self-directed learning, self-regulated learning, reading disabilities, language syllabus design, and learner autonomy. sbiryan@yahoo.com

Sema Karakelle is currently working as an Assistant Professor of Developmental Psychology Department at Istanbul University. Her research interests are metacognition, self-regulation in early childhood and development of theory of mind. semakarakelle@gmail.com

References

- Annevirta, T., Laakkonen, E., Kinnunen, R., & Vauras, M. (2007). Developmental dynamics of metacognitive knowledge and text comprehension skill in the first primary school years. *Metacognition and Learning, 2*, 21-39.
- Artzt, A. F., & Armour-Thomas, E. (1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. *Cognition and Instruction, 9*, 137
- Bingham, S. & Whitebread, D. (2008). Teachers supporting children’s self-regulation in conflict situations within an early years setting. In T. Paptheodorou & J. Moyles, (Eds.), *Learning Together in the Early Years: Explaining Relational Pedagogy* (pp. 91-102). London: Routledge.

- Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of Informational text: relations between processing and representation. *Discourse Processes* 25-1, 1-53.
- Cromley, J.G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning*, 1, 229-247.
- Cromley, J.G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning*, 1, 229-247.
- Demirel, M. (1995). Bilgilendirici metin türünün ve okuduğunu kavrama becerisinin altıncı sınıf öğrencilerinin öğrenme düzeyine etkisi [The effect of informational text type and reading comprehension skill on 6th graders learning achievement]. Unpublished doctoral dissertation. Hacettepe University, Ankara, Turkey.
- Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: how you test is what you get. *Metacognition and Learning*, 3(3), 189-206.
- Desoete, A., Roeyers, H., & De Clercq, A. (2003). Can off-line metacognition enhance mathematical problem solving? *Journal of Educational Psychology*, 95(1), 188-200.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Beverly Hills, CA: SAGE.
- Efklides, A. (2006). Metacognition and Affect: What can metacognitive experiences tell us about the learning process. *Educational Research Review*, 1(1), 3-14.
- Efklides, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema*, 21(1), 76-82.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised edition). Cambridge, MA: Bradford.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906-911.
- Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development*, 24, 15-23.
- Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, memory and Cognition*, 33(1), 238-244.
- Fortunata, I., Hecht, D., Carol, K. T. & Alvarez, L. (1991). Metacognition and problem solving. *The Arithmetic Teacher*, 39(4), 38-40.
- Hacker D. J., Bol L., Horgan D. D., & Rakow E. A. (2000). Text prediction and performance in a classroom context. *Journal of Educational Psychology*, 92, 160-170.
- Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and attributional style. *Metacognition and Learning*, 3, 101-121.
- Hannah, C. L., & Shore, B. M. (1995). Metacognition and high intellectual ability: Insights from the study of learning-disabled gifted students. *Gifted Child Quarterly*, 39, 95-109.
- Helms-Lorenz, M., & Jacobse, A. E. (2008). Metacognitive skills of the gifted from a cross-cultural perspective. In M. F. Shaughnessy, Veenman, M. V. J., & Kleyn-Kennedy, C.(Eds.), *Meta-Cognition: A recent review of research, theory and perspectives*. (pp. 3-43). New York, NY: Nova Science Publishers, Inc.
- Karakelle, S., & Saraç, S. (2007) Çocuklar için üst bilişsel farkındalık ölçeği (ÜBFÖ-Ç) A ve B formları geçerlik ve güvenilirlik çalışması [Validity and factor structure of Turkish versions of the metacognitive awareness inventory for children (Jr-MAI)- A ve B Forms]. *Turkish Psychological Articles*. 10(20), 87-103.

- Koriat, A. (2000). The feeling of knowing: Some meta-theoretical implications for consciousness and control. *Consciousness and Cognition, 9*, 149-171.
- Koriat, A. (2008). Are we frightened because we run away? Some evidence from metacognitive feelings. In B. Uttl, I. Ohta., & A. L. Siegenthaler (Eds.), *Memory and emotion: Interdisciplinary perspectives* (pp. 83-103). Malden, MA: Blackwell Publishing.
- Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R. A. (2006). Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 595-608.
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*, 147-162.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-99.
- Meijer, J. Veenman, M. V. J. & van Hout-Wolters, B. H. A. M. (2006). Metacognitive activities in text studying and problem solving: Development of taxonomy. *Educational Research and Evaluation, 12*(3), 209-237.
- Metcalfe, J. (2009). Metacognitive judgements and control of study. *Current Directions in Psychological Science, 18*(3), 159-163.
- Minnaert, A. & Janssen, P. J. (1999). The additive effect of regulatory activities on top of intelligence in relation to academic performance in higher education, *Learning and Instruction, 9*, 77-91.
- Mokhtari, K. & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*, 249-259.
- Myers, M., & Paris, S. G. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology, 70*(5), 680-690.
- Nietfeld, J., Cao, L. & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education, 74*(1), 7-28.
- Paris, S. G. & Jacobs, J. E. (1984). The benefits of informed instruction for children's reading awareness and comprehension skills. *Child Development, 55*, 2083-2093.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)* (Technical Report No. 91-B-004), Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.
- Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43-97). Lincoln, NE: The University of Nebraska Press.
- Pressley, M., & Afflerbach, P. (1995). Verbal protocols of reading: The nature of constructively responsive reading. Hillsdale, NJ: Erlbaum.
- Pressley, M., & Ghatala, E. S. (1989). Metacognitive benefits of taking a test for children and young adolescents. *Journal of Experimental Child Psychology, 47*, 430-450.
- Rast, P., & Zimprich, D. (2009). Age differences in the underconfidence-with-practice effect. *Experimental Aging Research, 35*(4), 400-431.
- Sapnas, K. G., & Zeller, R. A. (2002). Minimizing sample size when using exploratory factor analysis for measurement. *Journal of Nursing Measurement, 10*(2), 135-154.
- Schmitt, M. C. (1990). A questionnaire to measure children's awareness of strategic reading processes. *The Reading Teacher, 43*(7), 454-461.

- Schraw, G. & Dennison, R. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460-475.
- Schraw, G. (1997). The effect of generalized metacognitive knowledge on test performance and confidence judgements. *Journal of Study Education, 65*(2), 135-146.
- Schraw, G. (2009). Measuring Metacognitive Judgements. In D.J.Hacker, J.Dunlosky & A.C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp: 415-429). New York, NY: Routledge.
- Serra, M. J., & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory and Cognition, 31* (6), 1258-1266.
- Sperling, R. A., Howard, B. C. Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology, 27*, 51-79.
- Sperling, R. A., Howard, B. H., Staley, R., & DuBois, N. F. (2004). Metacognition and self regulated learning constructs. *Educational Research and Evaluation, 10*(2), 117-139.
- Tabachnick, B.G. & Fidell, L.S. (2007). *Using Multivariate Statistics*, 5th ed. Boston: Pearson Education, Inc.
- Thomas, K. F., & Barksdale-Ladd, M. A. (2000). Metacognitive processes: Teaching strategies in literacy education courses. *Reading Psychology, 21*, 67-84.
- van Kraayenoord, C. E., & Schneider, W. E. (1999). Reading achievement, metacognition, reading self-concept and interest: A study of German students in grades 3 and 4. *European Journal of Psychology of Education, 14*(3), 305-324.
- Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In B. Moschner & C. Artelt (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 75-97). Berlin: Waxmann.
- Veenman, M. V. J., & Beishuizen, J. J. (2004). Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraint. *Learning and Instruction, 14*, 621-640.
- Veenman, M. V. J., & Verheij, J. (2003). Technical students' metacognitive skills: Relating general vs. specific metacognitive skills to study success. *Learning and Individual Differences, 13*, 259-272.
- Veenman, M. V. J., Elshout, J. J., & Meijer, J. (1997). The generality vs. domain-specificity of metacognitive skills in novice learning across domains. *Learning and Instruction, 7*, 187-209.
- Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: Availability versus production deficiencies. *Anxiety, Stress and Coping, 13*, 391-412.
- Veenman, M. V. J., Kok, R., & Blöte, A. W. (2005). The relation between intellectual and metacognitive skills in early adolescence. *Instructional Science, 33*, 193-211.
- Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction, 14*, 89-109.
- Whitebread, D. & Coltman, P. (2010). Aspects of pedagogy supporting metacognition and self-regulation in mathematical learning of young children: evidence from an observational study. *ZDM, 42*(2), 163-178.
- Whitebread, D., Bingham, S., Grau, V., Pino Pasternak, D., & Sangster, C. (2007). Development of metacognition and self-regulated learning in young children: the role of collaborative and peer-assisted learning. *Journal of Cognitive Education and Psychology, 3*, 433-455.
- Whitebread, D., Coltman, P., Pino Pasternak, D., Sangster, C., Grau, V., Bingham, S., Almeqdad, Q., & Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulation in young children. *Metacognition and Learning, 4*(1), 63-85.

- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. E. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-regulation* (pp. 531–566). San Diego: Academic.
- Zimmerman B. J. , & Martinez-Pons M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology*, 80, 284-290.
- Zimmerman, B. J. & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating Grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, 82, 51–59.