
Assessing Metacognitive Skills in Computerized Learning Environments

11

Marcel V.J. Veenman

Abstract

In this chapter, metacognitive skills are considered to be an organized set of metacognitive self-instructions for the monitoring of and control over cognitive activity. These self-instructions can be represented as a production system of condition-action rules. For the assessment of metacognitive skills, however, these covert rules have to be inferred from overt learner behavior during task performance. In computerized learning tasks, on-line traces of learner activities can be unobtrusively stored in logfiles. Prerequisite to logfile assessment is the selection of relevant indicators of metacognitive learning activities on the basis of a rational task analysis, which indicators have to be validated against other on-line measures obtained with, for instance, thinking-aloud protocols. Such analyses of logfiles will allow for the assessment of metacognitive skills as an aptitude, that is, as a relatively stable repertoire of self-instructions. In order to further capture the dynamic change in metacognitive processes over time, progressive patterns of metacognitive activity can be identified in logged traces through time-series analysis. It is argued that the aptitude and dynamic approaches to assessing metacognitive skills are complementary to one another, rather than excluding each other.

Introduction

Metacognition is a relevant predictor of learning outcomes in traditional learning settings (Wang, Haertel, & Walberg, 1990) as well as in com-

puter-based learning environments (Veenman, 2008; Winters, Greene, & Costich, 2008). In conceptions of metacognition, a distinction is often made between knowledge of cognition and regulation of cognition (Brown, 1987; Schraw & Dennison, 1994). Metacognitive knowledge is declarative knowledge about the interplay between person characteristics, task characteristics, and strategy characteristics (Flavell, 1979). Having declarative metacognitive knowledge at one's disposal, however, does not guarantee that

M.V.J. Veenman (✉)

Department of Developmental and Educational
Psychology, Institute for Psychological Research,
Leiden University, Leiden, The Netherlands
e-mail: veenman@fsw.leidenuniv.nl

this knowledge is actually used for the regulation of learning behavior (Veenman, Van Hout-Wolters, & Afflerbach, 2006; Winne, 1996). Metacognitive knowledge may be incorrect or incomplete; the learner may fail to see the usefulness or applicability of that knowledge in a particular situation, or the learner may lack the skills for doing so.

Metacognitive skills refer to procedural knowledge that is required for the regulation of and control over one's learning behavior. Orientation, goal setting, planning, monitoring, checking, evaluation, and recapitulation are manifestations of those skills (Veenman, 2011). These skills directly affect learning behavior and, consequently, learning outcomes. Veenman (2008) estimated that metacognitive skillfulness account for about 40% of variance in learning outcomes for a broad range of tasks. Metacognitive skillfulness is regarded here as an aptitude, which is a relatively stable disposition for how the individual interacts with learning environments (Snow, 1989). This is not to say that metacognitive skills are entirely fixed. Learning experiences, instruction, and training may affect those skills (Pressley & Gaskins, 2006; Veenman, 2011). This chapter addresses issues related to the assessment of metacognitive skills in computer-based learning environments and, in particular, the necessity of validating these assessments through a multi-method approach.

Theoretical Framework

In an attempt to formulate a unifying theory of metacognition, Nelson (1996; Nelson & Narens, 1990) distinguished an "object level" from a "meta-level" in the cognitive system. At the cognitive level, lower-order cognitive activity takes place, usually referred to as *execution* processes. For instance, when solving a math problem, basic reading processes are needed for assessing the problem statement, and calculation processes are needed for producing the outcome. Higher-order *executive* processes of evaluation and planning at the meta-level govern the object level. Two flows

of information between both levels are postulated. Information about the state of the object level is conveyed to the meta-level through monitoring processes, while instructions from the meta-level are transmitted to the object level through control processes. Thus, if an error occurs on the object level, a metacognitive monitoring process will give notice of it to the meta-level, and control processes will be activated to resolve the problem.

Nelson's model essentially is a bottom-up process model. Anomalies in task performance trigger monitoring activities, which in turn activate control processes on the meta-level in order to restore cognitive processing at the object level. This model, however, does not clarify how monitoring processes themselves are activated (Dunlosky, 1998). Moreover, Nelson's model ignores the goal-directedness of human problem-solving and learning behavior as it does not allow for spontaneous activation of control processes without prior monitoring activity (Veenman, 2011). Koriath, Ma'ayan, and Nussinson (2006) have shown that causality in the relation between monitoring and control processes is bidirectional. Monitoring processes may elicit control processes, like Nelson emphasized, but control processes can also be activated without prior monitoring and, subsequently, elicit monitoring processes. The question, then, is how these control processes are activated if not by sheer coincidence.

Veenman (2011) extended Nelson's bottom-up model with a top-down approach. Metacognitive skills are perceived as an acquired program of self-instructions for the control over and the regulation of task performance. This program of self-instructions is activated whenever the learner is faced with a task that is familiar to the learner to a certain extent. Either the task has been practiced before or the task resembles another familiar task. These self-instructions can be represented as a production system of condition-action rules (Anderson, 1996; Winne, 2010). For instance, activating prior knowledge can be represented as: If you have read the task assignment, then retrieve all that you know about the topic from memory. Planning could be triggered by the rule: If you have set your goal, then design an action plan for

attaining that goal. Even self-induced, intentional monitoring is part of this production system: If you have executed a step from your action plan, then look out for errors in the executed step. This system of self-instructions is acquired through experience and training, much in the same way as the acquisition of cognitive skills (for more details, see Veenman, 2011). The more experienced a learner becomes, the more fine grained the condition-action rules will be with regard to, for instance, the selection of retrieval cues for memory search, the conversion from goal states to action plans, and the recognition of potential errors. In line with Nelson's model, self-instructions from the meta-level evoke various cognitive activities on the object level. However, self-instructions are self-induced, that is, they need not necessarily be triggered by a monitoring process of *anomalies* in task performance. In fact, the monitoring information flow in Nelson's model should be extended with the monitoring of conditions for activating self-instructions at the meta-level, although the latter is not necessarily a conscious process. Recognition of applicable conditions may also be automated to a certain extent in case self-instructions have become proficient metacognitive habits (Veenman et al., 2006). With reading, for instance, many monitoring processes run in the "background" of cognitive processes that are being executed. Proficient readers may not notice them, not even when thinking aloud. In this notion of self-instructions, the monitoring information flow represents the input to the production rule system at the meta-level. In the same vein, the control information flow represents the output of production rules.

What does this notion of self-instructions imply for the assessment of metacognitive skills? The aim of metacognitive assessment is to capture the learner's program of self-instructions at the meta-level. However, metacognitive skills that operate at the meta-level are not directly available for inspection (Veenman, 2011). The production system of self-instructions itself is covert and cannot be assessed, like the program lines of a compiled computer program that cannot be read. Verbalizations of the learner, however, can disclose the input and output of the production

system. Thus, the thinking-aloud method gives access to the monitoring and control information flow, and a production rule may be inferred from the relation between input and output information. For instance, we may hear a math learner say that the outcome of a calculation is odd. Yet, we *infer* from its cooccurrence with subsequent recalculation of the problem that a self-instruction for checking the outcomes must have been activated. Such inferences may be flawed, either because the input information is incomplete or because the output information is generated for another reason. When the math learner says "Let's do this again," this output information does not necessarily refer to recalculating the problem. Careful inspection of contingencies between monitoring and control information is warranted.

Most of the control information is gathered from overt operations on the object level. A task assignment is read, a sketch of the problem is drawn, a goal is written down, actions are taken step by step according to a plan, a dictionary is consulted for an unknown word, the flow of cognitive activity is interrupted for checking results, a recalculation is done, and conclusions are formulated. In fact, the execution of metacognitive skills draws heavily on lower-order cognitive processes (Veenman, 2011). From the perspective of the object level, one has to consider the context in which these cognitive processes occur in order to appraise their metacognitive origin. For instance, rereading is not a metacognitive activity as such, but it becomes a metacognitive activity if effortless reading is interrupted by the presence of a difficult word or a complex phrase. Thus, an inference process is required to identify specific cognitive activities at the object level and to tag them as "metacognitive activities." Unfortunately, this inference process is also prone to misinterpretation. Recalculation may be due to the metacognitive self-instruction of checking outcomes, but it may equally result from a learner's sloppiness in note taking. Observation techniques without concurrent thinking aloud or computer registrations of learner activities are more vulnerable to misinterpretations because they only have access to (metacognitive) activities at the object level.

Off-Line vs. Online Assessments

Generally, off-line methods for assessment of metacognitive skills are distinguished from online methods (Veenman, 2005). Off-line assessments concern the learners' self-reports that are gathered *prior to* or *after* task performance. Questionnaires (e.g., MSLQ, Pintrich & De Groot, 1990; MAI, Schraw & Dennison, 1994) and interviews (Zimmerman & Martinez-Pons, 1990) are off-line assessment methods that are frequently being used because they are relatively easy to administer. Off-line self-reports of metacognitive skills, however, suffer from validity problems. A first, fundamental problem concerns the off-line nature of self-reports, which requires learners to reconstruct their earlier performance. This reconstruction process might suffer from memory failure and distortions, especially if experiences from the past have to be retrieved (Veenman, 2011). The second validity problem is embedded in common questions about the relative frequency of certain activities ("How often do/did you...?"). In order to answer these questions, learners have to compare themselves to others (peers, parents, or teachers). The individual reference point chosen, however, may vary from one learner to another or even within a learner from one question to another (Veenman, Prins, & Verheij, 2003). Variation in reference points chosen by learners may yield disparate data. It is much like measuring the temperature with differently scaled thermometers, however, without being able to rescale measurements. Moreover, some learners may produce socially desirable answers.

Online assessments are obtained *during* task performance, that is, they are based on actual performance of the learner. Typical online assessments include observational methods (Whitebread et al., 2009), the analysis of thinking-aloud protocols (Azevedo, Greene, & Moos, 2007; Pressley & Afflerbach, 1995; Veenman, Elshout, & Meijer, 1997), and eye-movement registration (Kinnunen & Vauras, 1995). The essential difference between off-line and online methods is that off-line measures merely rely on learner self-reports, whereas online measures pertain to the coding of actual

learner behavior on externally defined criteria by external agencies, such as "blind" judges and observers (Veenman, 2011). The use of a standardized coding system circumvents the validity problems mentioned before. Online assessments also have their limitations. Thinking aloud may not always yield complete protocols, for instance, when processes are highly automated or, conversely, when the task is extremely difficult (Ericsson & Simon, 1993). Observed behavior needs to be interpreted by observers whenever the learner fails to express the reasons for his/her conduct (Veenman, 2011). Similarly, the registration of eye movements only captures the motor activities of the eyes. The meaning of these overt activities is subject to interpretation for which the coding system should provide perspicuous standards.

Research with multi-method designs has shown that off-line measures hardly correspond to online measures. Correlations between off-line and online measures are invariably low ($r=0.15$ on the average; Bannert & Mengelkamp, 2008; Cromley & Azevedo, 2006; Veenman, 2005, 2011; Veenman et al., 2003), and qualitative analyses show that off-line self-reports do not converge with specific online behaviors (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; Winne & Jamieson-Noel, 2002). Apparently, learners do not do what they prospectively say they will do nor do they accurately recollect what they have recently done. Moreover, correlations among off-line measures are often low to moderate, whereas correlations among online measures are moderate to high (Cromley & Azevedo, 2006; Veenman, 2005). Obviously, off-line methods yield rather diverging results, while online methods converge in their assessments of metacognitive skills. Finally, the external validity of assessment methods should be considered (Veenman, 2007). Online assessments are strong predictors of learning outcomes, contrary to off-line assessments. In a review study, Veenman (2005) found that correlations with learning performance range from slightly negative to 0.36 for off-line measures and from 0.45 to 0.90 for online measures. In conclusion, off-line measures suffer from low convergent validity and low external validity,

which makes an argument for resorting to online assessment of metacognitive skills (Veenman, 2007). Yet, a majority of studies rely on off-line self-reports for the assessment of metacognition (Dinsmore, Alexander, & Loughlin, 2008; Veenman, 2005), including studies with computer-based learning environments (Gress, Fior, Hadwin, & Winne, 2010; Winters et al., 2008).

Logfile Assessments

Thinking aloud and observation are time-consuming methods because they have to be administered on an individual basis. With the emergence of computer-based learning environments, the online method of tracing metacognitive behaviors of learners in computer logfiles has become available (Greene, & Azevedo, 2010; Hadwin et al., 2007; Kunz, Drewniak, & Schott, 1992; Veenman, Elshout, & Groen, 1993; Veenman, Wilhelm, & Beishuizen, 2004; Winne, 2010). Obviously, the nature of the task should allow for a computerized version, or otherwise it would impair the ecological validity of assessments. The advantage of logfile assessment is that the method is minimally intrusive and that it can be administered to large groups at the same time (Aleven, Roll, McLaren, & Koedinger, 2010; Azevedo, Moos, Johnson, & Chauncey, 2010; Veenman et al., 2006; Winne, 2010). Typically, a logfile contains traces of the learner's overt cognitive activities during task performance on the computer. The frequencies of certain key presses, button pushes, object manipulations, link and screen selections, scrolling, and menu clicks are registered along with time indications. Logfiles do not contain the learner's metacognitive deliberations for enacting those activities, since prompting learners to type in their thoughts would interfere with spontaneous metacognitive processing. Basically, the concrete activities registered in a logfile represent rather raw materials on a low cognitive level, also referred to as "events" (Azevedo et al., 2010; Winne, 2010). In order to lift logfile analysis to a metacognitive level, two steps need to be taken in order to select and validate relevant indicators of metacognition (Veenman, 2007).

A first step in logfile analysis concerns the selection of which cognitive activity may be consequential to metacognitive regulation. This selection of potential indicators of metacognitive skills should be based on a rational analysis of the task at hand, knowledge of the metacognition literature, and common sense. For instance, pushing a particular button at a critical moment in the course of task performance may be such an indicator. The outcome of this selection process, however, is not always entirely successful. Some activities that initially appear to be metacognitive by nature may turn out to be non-metacognitive after all. Hence, a second step is to validate these potential logfile measures with concurrent online assessments, such as think-aloud protocols or systematical observation. This multi-method approach is prerequisite for establishing a firm set of adequate logfile indicators of metacognitive skillfulness (Veenman, 2007; Winters et al., 2008). Selection and validation of indicators need to be done prior to logfile assessments if the coding of learner activities in logfiles is automated. Otherwise, logfiles have to be coded by hand afterwards. Three empirical studies may elucidate the necessity of this two-step procedure.

Veenman and colleagues (1993) assessed the metacognitive skills from logfiles of 40 participants who were either thinking aloud or working silently in a computer-simulated Heat Lab. Participants, novices in the domain of physics, were required to discover principles of calorimetry by designing their own experiments. Several objects of different weights (100 g, 200 g, 1 kg) and materials (gold, copper, glass) could be heated on a burner. The amount of heat transferred to an object was regulated with a time switch and could be read off a joules-meter. Temperature was measured by attaching a thermometer to an object. Thus, the virtual laboratory contained the required means for examining the relationship between heat and temperature depending on weight and material. All activities in Heat Lab were logged. In order to determine which of these activities could be labeled as representing metacognitive skillfulness, a reference group with a similar background was included from an earlier study with Heat Lab

(Veenman et al., 1993). Thinking-aloud protocols of this reference group had been analyzed on the quality of metacognitive skillfulness (i.e., on indications of task orientation, goal setting, planning, monitoring, evaluation, recapitulation, and reflection). Their logfiles were coded on potential positive indicators of metacognitive orientation (frequency of rereading the task assignment and frequency of asking for help with lab operations), positive indicators of planning (frequency of switching on the burner for starting a new experiment, frequency of object manipulations, and the number of unique objects used), as well as negative indicators of planning and monitoring (frequency of *not* measuring either the initial temperature or the final temperature). Although the selection of these indicators was based on a rational task analysis, only three logfile measures appeared to be substantially related to thinking-aloud measures. Frequency of switching the burner on, frequency of *not* measuring the initial temperature, and frequency of *not* measuring the final temperature correlated 0.40, -0.40, and -0.37 with the thinking-aloud measures, while correlations for the other logfile measures were low. Regression analysis confirmed that these three logfile measures each contributed to the prediction of the thinking-aloud measure, whereas the others did not. A composite score of these three logfile measures correlated 0.62 with the thinking-aloud measures. Using the same procedure for obtaining a composite score from logfiles in the main experiment, Veenman and colleagues (1993) showed that participants who were thinking aloud did not differ in metacognitive skillfulness from those who worked silently, $F(1,38)=0.02$. Thinking aloud did not affect metacognitive processes, although it slowed down those processes a bit (cf. Ericsson & Simon, 1993).

In another study, Veenman and colleagues (2004) assessed metacognitive skillfulness from the logfiles of 113 children and adolescents in the age of 9–22 years, who performed four computer-simulated, inductive-learning tasks. Participants completed two biology tasks (a plant-growing task and a food task) as well as two geography tasks (one about the conservation of otter habitats, the other about ageing). In each

task, five independent variables with discrete levels (either two or three levels) could be varied, and their effects on the dependent variable could be inspected. The model underlying the relations between the independent and the dependent variables was identical in each task; two independent variables interacted with one another; one variable had a nonlinear effect, and two variables were irrelevant. Each task model corresponded to plausible real-life phenomena. Figure 11.1 shows the interface of the plant-growing task as an example. The task was to find out how different independent variables affected plant growth. Independent variables were (1) giving water, either once or twice a week; (2) using an insecticide or not; (3) putting dead plant leaves in the flowerpot or not; (4) placing the plant either indoors, on a balcony, or in a greenhouse; and (5) size of the flower pot, either large or small. Distinct levels of plant growth as a dependent variable were 5, 10, 15, 20, and 25 cm. Variable 4 had a nonlinear effect, meaning that growing the plant indoors resulted in 5 cm less growth, relative to a balcony or greenhouse. Variable 2 and 3 did not affect plant growth at all. Variable 1 and 5 interacted, as giving water once or twice a week did not matter for a large pot, but it did matter when a small flowerpot was used. In that case, giving water twice a week would reduce plant growth, while giving water once a week would increase growth, relative to growth in the large flowerpot. Within each task, participants performed a series of “experiments.” Such an experiment consisted of choosing a value for each of the independent variables, predicting the plant growth as a result of these values, and asking the computer for the actual plant growth. Results of earlier experiments could be inspected by scrolling through the result window at the right side in Fig. 11.1.

During the food task, participants had to find out how eating and drinking habits affected the health status of an imaginary person, called Hans. Independent variables were the consumption of fat, carbohydrates, alcohol, albumen, and supplementary vitamins. In the other task, the relevance of factors affecting the extinction of otters had to be investigated. Independent variables were extra

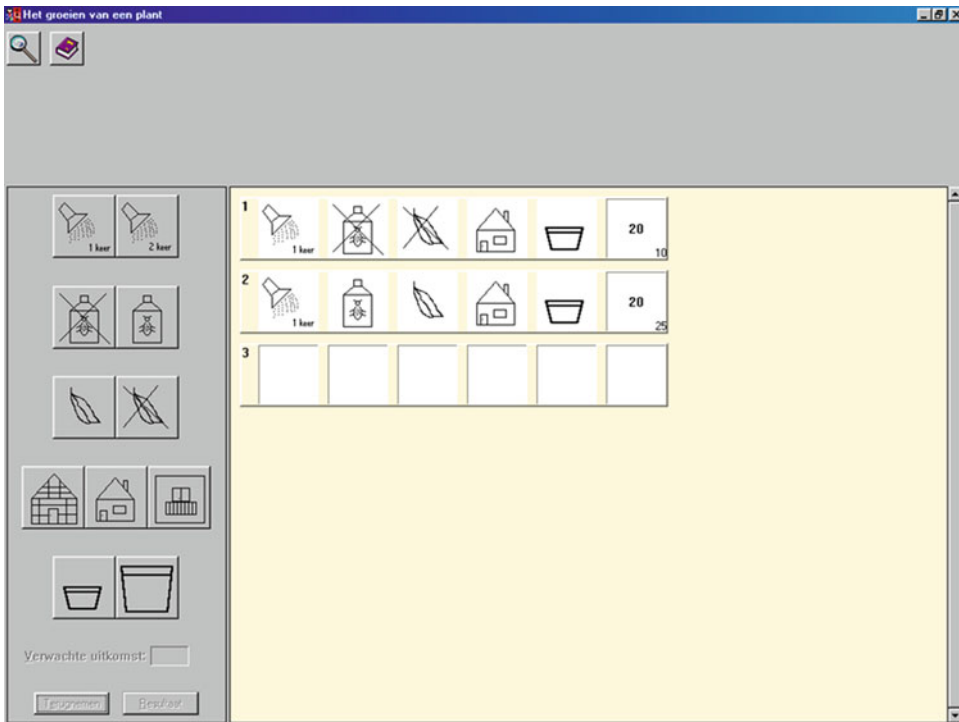


Fig. 11.1 Interface of the plant-growing task (Veenman et al., 2004)

food provision or not, environmental pollution, natural habitat, media exposure, and closing otter areas to the public or not. In the population-ageing task, independent variables that could affect the ageing rate of a population were state of the economy, quality of the educational system, means of living, climate, and general safety. In all cases, two variables interacted, one variable had a nonlinear effect, and two variables were irrelevant to the dependent variable.

The computer program automatically recoded learner activities in the logfiles of each task into several potential indicators of metacognitive skillfulness. Logfile measures included the number of (unique) experiments conducted, the mean number of variables changed per experiment, frequency of scrolling activities, frequency of variable selection activities, the prediction-error rate (mean distance between predicted and actual outcome), and time on task, among others (see Wilhelm, Beishuizen, & Van Rijn, 2005). As participants were required to think aloud during all tasks, logfile measures could be validated against

thinking-aloud data. Two judges separately rated 10% of the plant-growing-task protocols and 5% of the otter-task protocols on the quality of metacognitive skillfulness. Protocols were judged on the quality of orientation activities (elaborateness of hypotheses generated before each experiment), systematical behavior (planning a sequence of experiments and avoiding unsystematic variations between subsequent experiments), evaluation (detection and correction of mistakes), and elaboration (drawing conclusions, relating outcomes of experiments, generating explanations, and recapitulating). From the logfile measures, only the mean number of variables changed per experiment, and the frequency of scrolling appeared to be substantially correlated to the thinking-aloud scores. The mean number of variables changed per experiment (VOTAT; Chen & Klahr, 1999) was a negative indicator of think-aloud metacognition. Varying more than one variable at a time represents poor planning behavior (Veenman et al., 1997) and lack of experimental control (Glaser, Schauble, Raghavan, & Zeitz,

1992). Frequency of scrolling back to earlier experiments, on the other hand, was a positive indicator of think-aloud metacognition. Participants use scrolling to check earlier experimental configurations or to relate outcomes of experiments. Scores on both measures were standardized and the sign of the negative indicator was inverted. Composite scores of these two logfile measures correlated 0.85 and 0.84 with the thinking-aloud data of the plant-growing task and the other task, respectively. Veenman and colleagues (2004) used the composite scores of logfile measures to show that the metacognitive skills of learners develop with age. The mean composite scores of the four age groups (9, 12, 14, and 22 year) revealed a steep linear increment with age, $F(3,109)=38.60$, $p<0.001$. Moreover, composite scores correlated 0.74 with an overall measure of learning performance.

In a recent, unpublished study, Veenman, Van Haaren, and Rens used an adapted version of the plant-growing task to assess the metacognitive skills of gifted secondary-school students. Task complexity was increased to meet the intellectual level of the target group of gifted students. Numerical relations between the variables were made more complex, and a second interaction effect was included: Using insecticides with dead leaves in the pot reduced the growth of the plant, while leaving out any of the two did not affect plant growth. Due to changes in both task settings and target group compared to the original study, a pilot study with five gifted learners needed to be carried out to validate logfile measures once more. Think-aloud and logfile data were gathered according to the procedures of Veenman and colleagues (2004). A new posttest with multiple-choice and open-ended questions about the effects of the five independent variables on plant growth was also administered. As expected, VOTAT (converted to positive scores) and the frequency of scrolling activities correlated respectively 0.68 and 0.58 with think-aloud metacognition. However, this time the number of unique experiments, corrected for the total number of experiments, correlated 0.56 with think-aloud metacognition. The number of unique experiments represents coverage of the

problem space, consisting of maximal 48 possible experiments. Composite scores of these three logfile measures correlated 0.96 ($p<0.01$) with think-aloud metacognition and 0.90 ($p<0.05$) with posttest learning outcomes.

The first two studies show that a selection of logfile indicators based on a rational task analysis is fallible. Validation of potential indicators is necessary to sift out irrelevant, non-metacognitive activities. Moreover, the third study reveals that additional validation is required when task conditions or participant samples are altered. These studies further show that a limited set of logfile measures may adequately represent a broader range of metacognitive skills assessed from thinking-aloud protocols. Veenman and colleagues (2004) asserted that metacognitive skills during various phases of task performance are highly interdependent. Good orientation leads to good planning and systematical behavior, which in turn allows for more monitoring and evaluative control. This interdependency of metacognitive skills (with intercorrelations of about 0.90; Veenman, 1993) accounts for why a limited set of indicators may adequately represent broad metacognition.

Patterns of Activity in Logfile Assessments

Logfiles assessments often merely capture the quantity of metacognitive activities and not the quality of those activities (Winters et al., 2008). Plain rereading of task assignments, for instance, is not the same as rereading the task assignment consequential to monitoring the understanding of the task. The latter is more goal oriented. One way to access quality is to detect meaningful patterns in the sequence of activities or events. Transition analysis is used to analyze trace data on the sequence and transitions of events (Azevedo et al., 2010; Hadwin et al., 2007). All frequencies of transitions from one event to another are entered in a matrix of all possible events. Inspection of this matrix yields information about the regularity of certain transitions (density) and about the exclusivity of transition

starting points (centrality). Transition analysis may be done on group level as well as on the individual level. In the same vein, Biswas, Hogeong, Kinnebrew, Sulcer, and Roscoe (2010) used a technique of hidden Markov models to detect probability patterns of transitions between (metacognitive) activities over time. Such techniques allow us to detect patterns of contingent events, rather than registering single, isolated ones (Winne, 2010). The metacognitive nature of these patterns, however, remains to be inferred by the researcher.

Researchers in self-regulated learning stress the dynamic nature of metacognitive processes (Azevedo et al., 2010; Greene & Azevedo, 2010; Winne, 2010). Strategy choices and frequency of activities may change over time in interaction with the learning environment. Time-series analysis is a technique for assessing changes in metacognitive functioning. For time-series analyses, either the task is subdivided in distinguishable learning episodes, or a series of highly similar tasks is presented. Repeated assessments over time are analyzed. In the unpublished study of Veenman, Van Haaren, and Rens, eventually 153 students from preuniversity secondary education performed both the plant-growing task and the ageing task in randomized order. Logfile measures were analyzed by means of repeated-measures ANOVA with task order as between-subjects factor. Results show that the total number of experiments, $F(1,151)=36.84$, $p<0.001$, converted VOTAT, $F(1,151)=54.09$, $p<0.001$, and the number of unique experiments, $F(1,151)=26.57$, $p<0.001$, increased between task 1 and task 2, while the scrolling frequency, $F(1,151)=22.80$, $p<0.001$, decreased. Participants became more active and showed more experimental control over time, at the cost of scrolling activities. Perhaps, referring back to previous experiments became less compulsory due to the enhanced experimental control.

Elshout, Veenman, and Van Hell (1993) used time-series analysis to study help-seeking behavior in a computerized learning-by-doing environment. Novice and advanced learners in physics learned to solve a series of 20 complex thermodynamics problems about the relation between volume,

pressure, and temperature with the option of asking for help from the computer program. The help facility offered a sequence of steps that would lead the learner through an orientation phase, an execution phase, and an evaluation phase of the problem-solving process. Participants were free to choose a type of help: clue (hint about one specific step), one step (working out of one specific step), student performed (all subsequent steps, but learner executed), or computer demonstrated (working out of all steps, demonstrated by the program). Traces of help requests were logged, while metacognitive skillfulness was assessed from think-aloud protocols. Analysis over the series of 20 problems revealed that metacognitively poor novices preferred the quick and dirty way out by choosing one-step formula with direct access to a working out of the appropriate formula (cf. Alevén et al., 2010). Help requests of metacognitively skilled novices, on the other hand, shifted from merely execution help to orientation help over the 20 problems, thereby matching the help-seeking behavior of advanced learners in the end. These two studies show that time-series analysis of logged traces may capture patterns of change in metacognitive functioning.

Discussion

In the introduction of this chapter, metacognitive skillfulness was defined as an aptitude. Recently, Winne (2010) argued against such an aptitude approach because self-regulation is a dynamic process that unfolds in the course of learning. Self-regulatory processes change in nature and frequency as learning progresses. According to Winne, aptitude measures do not capture the dynamic nature of self-regulation, contrary to computer traces of events that allow for a fine-grained analysis of processes over time. The construct of metacognitive skillfulness is an aptitude indeed, because it represents the availability of self-instructions in learning situations. Assessments of metacognitive skills as an aptitude would provide a static measure of the amount and quality of available skills

(Winne, 2010). Yet, both positions of metacognitive skills as aptitude and as dynamic processes are equally tenable, provided that metacognitive skills are assessed with behavior measures. Studies have shown that learners bring along a rather stable, general repertoire of metacognitive skills when entering various new learning situations (Veenman, 2011). The deployment of this general repertoire, however, must be adapted to task demands and other contextual factors during the learning process, as shown by the studies with time-series analysis. Metacognitive skills are gradually tailored to the task at hand because production rules become more specialized and sensitive to task constraints. Thus, any learning experience may alter the repertoire of production rules for metacognitive self-instruction. Veenman (1993) postulated that a separate set of task- or domain-specific production rules is generated during the acquisition of expertise, alongside the general production rules that serve as default repertoire for novel learning situations. Even general production rules are subject to change due to experience and training, yet at a slow pace (Veenman et al., 2006). Therefore, the notion of metacognitive skills as self-instructions does not preclude a peaceful coexistence of aptitude and dynamic change in learning.

There is ample evidence that online assessments are more valid than off-line assessments of metacognitive skills. Nevertheless, all online assessments make inferences about metacognitive self-instructions, albeit to a different extent. The think-aloud method is a powerful tool for assessing monitoring and control information flows. Yet, protocols may be incomplete and researchers have to fill in the gaps by making inferences about relations between both information flows. The same is true for observations that include the learner's verbalizations. More far-reaching inferences need to be made for observations without verbalization, eye-movement registration, and logfile analysis as these methods only access information about concrete, overt behaviors on the object level of Nelson's model. For two contingent events, the researcher has to infer the

causal relation between the two events and their metacognitive nature. First, one needs to infer that the first event represents the condition part of a production rule. Next, one needs to infer that the second event corresponds to the action part of the *same* production rule. Finally, the metacognitive function of the entire production rule has to be inferred. Contingencies in time may offer a plausible but not sufficient reason for making these inferences (cf. Winne, 2010). A further complication is that the conditions for evoking a control event may not become manifest in trace logs, either because the conditions of a production rule are activated by mental operations that are not accessible with trace data or because the tracing system is not sensitive to a particular event. Here is a major challenge that researchers of trace data in computer-based learning environments are facing: extracting an appropriately contextualized (i.e., neither overly general nor overly specific) set of conditions from multiple data points.

Logfile assessment is an unobtrusive method for gaining access to events in detail on the object level, which assessments can be done on a large scale and over extended periods of time. Logfile analysis allows for different levels of granularity in assessment, ranging from tracing the occurrence of separate events to detecting patterns of contingent events. Although scarcely out of the egg, tracing events can be used for attuning feedback and scaffolding of metacognitive functioning to individual needs (Alevin et al., 2010; Azevedo et al., 2010; Gress et al., 2010) and for verifying that these interventions have been successful (Veenman, 2007). However, validation of logfile events with other online assessments is prerequisite to making justified inferences about the metacognitive nature of those events. Ultimate assessments would include different online methods rendering data that are aligned in time and produce converging results. Like a converging lens that directs rays of light to a focal point, even the focal distance may change due to learning experiences. Unfortunately, multi-method research in metacognition is scarce so far (Veenman, 2011; Veenman et al., 2006). Metacognition researchers should sharpen their lenses.

References

- Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2010). Automated, unobtrusive, action-by-action assessment of self-regulation during learning with an intelligent tutoring system. *Educational Psychologist, 45*, 224–233.
- Anderson, J. R. (1996). *The architecture of cognition*. Mahwah, NJ: Erlbaum.
- Azevedo, R., Greene, J. A., & Moos, D. C. (2007). The effect of a human agent's external regulation upon college students' hypermedia learning. *Metacognition and Learning, 2*, 67–87.
- Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist, 45*, 210–223.
- Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalization method affect learning? *Metacognition and Learning, 3*, 39–58.
- Biswas, G., Hogyeong, J., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology Enhanced Learning, 5*, 123–152.
- Brown, A. (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, NJ: Erlbaum.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*, 1098–1120.
- Cromley, J. G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning, 1*, 229–247.
- Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*, 391–409.
- Dunlosky, J. (1998). Epilogue. Linking metacognitive theories to education. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 367–381). Mahwah, NJ: Erlbaum.
- Elshout, J. J., Veenman, M. V. J., & Van Hell, A. G. (1993). Using the computer as a help tool during learning by doing. *Computers in Education, 21*, 115–122.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*. Cambridge: MIT Press.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906–911.
- Glaser, R., Schauble, L., Raghavan, K., & Zeitz, C. (1992). Scientific reasoning across different domains. In E. de Corte, M. C. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving* (NATO ASI series F, Vol. 84, pp. 345–371). Heidelberg: Springer.
- Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist, 45*, 203–209.
- Gress, C. L. Z., Fior, M., Hadwin, A. F., & Winne, P. H. (2010). Measurement and assessment in computer-supported collaborative learning. *Computers in Human Behavior, 26*, 806–814.
- Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning, 2*, 107–124.
- Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction, 5*, 143–165.
- Koriath, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*, 36–69.
- Kunz, G. C., Drewniak, U., & Schott, F. (1992). On-line and off-line assessment of self-regulation in learning from instructional text. *Learning and Instruction, 2*, 287–301.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*, 102–116.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). New York: Academic Press.
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33–40.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- Pressley, M., & Gaskins, I. (2006). Metacognitive competent reading is constructively responsive reading: How can such reading be developed in students? *Metacognition and Learning, 1*, 99–113.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460–475.
- Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences*.

- Advances in theory and research* (pp. 13–59). New York: Freeman.
- Veenman, M. V. J. (1993). *Intellectual ability and metacognitive skill: determinants of discovery learning in computerized learning environments*. Doctoral dissertation. Amsterdam: University of Amsterdam.
- Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und metakognition: Implikationen für forschung und praxis* (pp. 75–97). Berlin: Waxmann.
- Veenman, M. V. J. (2007). The assessment and instruction of self-regulation in computer-based environments: A discussion. *Metacognition and Learning*, 2, 177–183.
- Veenman, M. V. J. (2008). Giftedness: Predicting the speed of expertise acquisition by intellectual ability and metacognitive skillfulness of novices. In M. F. Shaughnessy, M. V. J. Veenman, & C. Kleyn-Kennedy (Eds.), *Meta-cognition: A recent review of research, theory, and perspectives* (pp. 207–220). Hauppauge: Nova.
- Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197–218). New York: Routledge.
- Veenman, M. V. J., Elshout, J. J., & Groen, M. G. M. (1993). Thinking aloud: Does it affect regulatory processes in learning. *Tijdschrift voor Onderwijsresearch*, 18, 322–330.
- 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., Prins, F. J., & Verheij, J. (2003). Learning styles: Self-reports versus thinking-aloud measures. *British Journal of Educational Psychology*, 73, 357–372.
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1, 3–14.
- 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.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *The Journal of Educational Research*, 84, 30–43.
- Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., & Bingham, S. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning*, 4, 63–85.
- Wilhelm, P., Beishuizen, J. J., & Van Rijn, H. (2005). Studying inquiry learning with FILE. *Computers in Human Behavior*, 21, 933–943.
- Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences*, 8, 327–353.
- Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, 45, 267–276.
- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibrations of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27, 551–572.
- Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning with computer-based learning environments: A critical analysis. *Educational Psychology Review*, 20, 429–444.
- 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.