Towards efficient measurement of metacognition in mathematical problem solving

Annemieke E. Jacobse · Egbert G. Harskamp

Received: 16 June 2011 / Accepted: 25 April 2012 /

Published online: 26 May 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Metacognitive monitoring and regulation play an essential role in mathematical problem solving. Therefore, it is important for researchers and practitioners to assess students' metacognition. One proven valid, but time consuming, method to assess metacognition is by using think-aloud protocols. Although valuable, practical drawbacks of this method necessitate a search for more convenient measurement instruments. Less valid methods that are easy to use are self-report questionnaires on metacognitive activities. In an empirical study in grade five ($n=39$), the accuracy of students' performance judgments and problem visualizations are combined into a new instrument for the assessment of metacognition in word problem solving. The instrument was administered to groups of students. The predictive validity of this instrument in problem solving is compared to a well-known think-aloud measure and a self-report questionnaire. The results first indicate that the questionnaire has no relationship with word problem solving performance, nor the other two instruments. Further analyses show that the new instrument does overlap with the think-aloud measure and both predict problem solving. But, both instruments also have their own unique contribution to predicting word problem solving. The results are discussed and recommendations are made to further complete the practical measurement instrument.

Keywords Measurement · Metacognition · Monitoring · Questionnaire · Performance judgments · Mathematics

Introduction

In the past years, metacognition has been recognized as one of the most relevant predictors of accomplishing complex learning tasks (Van der Stel and Veenman 2010; Dignath and Buttner 2008). Metacognition refers to meta-level knowledge and mental actions used to steer cognitive processes. In our study, we adopt the view of applied metacognition as consisting of metacognitive monitoring and regulation (Efklides 2006; Nelson 1996).

A. E. Jacobse (✉) · E. G. Harskamp

GLON, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, The Netherlands
e-mail: a.e.jacobse@rug.nl

Metacognitive regulation refers to mental activities used to regulate cognitive strategies to solve a problem (Brown and DeLoache 1978). For instance, when taking a note, the decision to do so is metacognitive, while the writing in itself is cognitive. Metacognitive monitoring refers to students' ongoing control over these learning processes. Monitoring can be used to identify problems and to modify learning behavior when needed (Desoete 2008). A large number of studies have already been undertaken to show that through metacognitive training, students' ability to solve mathematics problems improves (i.e. Jacobse and Harskamp 2009). For researchers, as well as teachers, it is important to have an adequate instrument to measure students' metacognition in order to analyze the relationship between growth in metacognition and growth in achievement. However, how to measure metacognition efficiently is still a problem. This problem has been at the heart of a great deal of scientific debate about which instruments are most suitable (Schellings and Van Hout-Wolters 2011).

One proven effective method to get insight into students' metacognition is asking them to verbalize their thoughts while working on a task. The verbalized thoughts are recorded and fully transcribed or judged by means of systematical observation (Veenman et al. 2005). This measurement technique is called think-aloud. Think-aloud protocols provide rich information on the metacognitive processes used during a learning task and are powerful predictors of test performance (Schraw 2010; Veenman 2005). A major strength of the use of think-aloud protocols, is that information about metacognitive behavior is collected directly when it is executed. This makes the information less vulnerable to students' memory distortions. Besides, students do not have to judge the appropriateness of their learning processes themselves (Veenman 2011b). Although sometimes slowing learning down, when executed correctly think-alouds do not impair students' learning performance (Bannert and Mengelkamp 2008; Fox et al. 2011). However, besides these positive characteristics, there is a major drawback of the method: Gathering and scoring the data of individual students' think-aloud protocols is a complex and time-consuming process which makes this measure inappropriate for test assistants or teachers who lack experience using the method, and for application in larger samples of students (Azevedo et al. 2010; Schellings 2011). Thus, when using this theoretically grounded measure, it tends to conflict with some more practical constraints of time and effort. Balancing theoretical and practical issues in the measurement of metacognition is a particularly challenging issue (McNamara 2011). In order to make measurements of metacognition more practical, it is important to explore the use of other instruments.

Researchers have already proposed several alternative measurement instruments to assess metacognition in a more practical manner, such as various self-report questionnaires. However, few of these instruments show convergence with think-aloud measures as predictors of performance. In this study pros and cons of alternative instruments are discussed that may substitute think-aloud protocol analysis. Alternative instruments which are shown in the literature to be valid indicators of metacognition are combined into a new measurement instrument. This measurement instrument can be collected in a paper-and-pencil format for larger groups of students which makes it notably easier to use than think-aloud measures. Explorative analyses comparing the new instrument with think-aloud scores are performed in a grade 5 sample, eventually aiming at the development of a more practical measurement instrument of students' metacognition in mathematics.

Theoretical framework

When measuring metacognition, it is important to note that metacognition probably is quite domain-specific (Veenman and Spaans 2005). The regulation of cognitive activities useful in

one domain (e.g. making a summary when reading) may not be directly transferable to another domain (e.g. solving a math problem). It is thus advisable to be specific about the context in which metacognition is measured (McNamara 2011). One of the domains in which metacognition is a key variable predicting learning performance is the domain of mathematical problem solving (Desoete and Veenman 2006; Desoete 2009; Fuchs et al. 2010; Harskamp and Suhre 2007). In this domain, metacognition is used to monitor solution processes and to regulate the problem solving episodes of analyzing and exploring a task, making a solution plan, implementing the plan and verifying the answer (Schoenfeld 1992). Such metacognitive processes can be measured off-line or on-line of the learning process. Online methodologies capture any activity that occurs during processing, whereas offline methods capture any activity that happens either before or after processing (Azevedo et al. 2010). Metacognition measured on-line of the learning process typically explains about 37 percent of the variance in learning (Veenman et al. 2006).

One of the most frequently used categories of off-line measures is self-report questionnaires in which students are asked to report on their own metacognition. Some examples of frequently used questionnaires are the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich and De Groot 1990), the Learning and Study Strategies Inventory (LASSI; Weinstein et al. 1988) and the Metacognitive Awareness Inventory (MAI; Schraw and Dennison 1994). These questionnaires typically contain quite general statements about metacognitive monitoring or regulation for which the student is asked to rate the degree to which the statement applies. Statements are used such as: "Before I begin studying I think about the things I will need to do to learn" or "I ask myself questions to make sure I know the material I have been studying" (Pintrich and De Groot 1990). One notable practical advantage of using questionnaires is that they can easily be administered on a large scale (Schellings and Van Hout-Wolters 2011). Besides, various studies in mathematical problem solving have shown the practicality and good internal consistency of self-report questionnaires (Kramarski and Gutman 2006; Mevarech and Amrany 2008). However, off-line measures do not measure learners' ongoing metacognitive behavior during task processing because they are collected before or after the student processes a learning task (Greene and Azevedo 2010). This causes some severe problems. Firstly, the fact that self-report questionnaires are collected separate from the learning task means that students have to retrieve earlier processes and performance from their long term memory. Self-report questionnaires thus are susceptible to memory distortion issues (McNamara 2011; Schellings 2011; Veenman 2011b). Secondly, students can differ in their frame of reference as to which situations they have in mind when answering the questions and interpreting the scales (McNamara 2011; Schellings 2011). Thirdly, the way students answer self-report questionnaires may be biased by triggers in the questions which prompt them to wrongly label their own behavior or by social desirability (Cromley and Azevedo 2011; Veenman 2011a). Therefore, students are typically quite inaccurate in reporting their own metacognitive behavior. Although self-report questionnaires are mostly designed to measure metacognitive regulation, they do not seem to be representative of what students actually do. This is illustrated by the fact that students' self-reported metacognitive behavior has found to be a poor predictor of performance. In a review of 21 studies using self-report questionnaires, the mean variance explained by metacognition in learning performance did not exceed 3 % ($r=0.17$) (Veenman and Van Hout-Wolters 2002). Additionally, some studies have shown the convergent validity between different questionnaires, theoretically measuring the same metacognitive processes, to be quite modest (Muis et al. 2007; Sperling et al. 2002). As some authors argue, off-line, generally formulated metacognitive questionnaires may be more adequate to assess metacognitive knowledge as opposed to metacognition applied during the learning process (Desoete 2007; Greene and Azevedo 2010).

On-line measures on the other hand have the advantage of measuring metacognition concurrent with the learning behavior, thus giving more insight in the actual use of metacognition affecting learning behavior. One way to infer on-line information about students' metacognition, apart from using think-aloud protocols as discussed before, is to assess the actions or observable occurrences of events that a student performs such as drawing schemes, taking notes or clicking a button (Winne and Perry 2000). Although in this case no direct information is gathered about the meta-level processes preceding the event, certain characteristics of the actions can be used to infer this information. In mathematical problem solving, an important cognitive action is making a drawing of the problem situation. Few students in elementary school use this strategy spontaneously. However, instructing students to make a drawing, can clarify how they think about solving a word problem (Van Essen and Hamaker 1990). Students' problem visualizations in a drawing can be either schematic or pictorial. In schematic visualizations the structural relationships between variables in a problem are provided in a sketch, diagram or schema. In pictorial visualizations the elements in a problem are depicted without any relevant relationships between the elements. Pictorial visualizations show a student does not yet know how to explore the problem towards a useful solution, thus indicating low metacognitive regulation. Visualizations that schematize problem situations on the other hand, are an expression of sophisticated metacognitive regulation in mathematical problem solving, especially giving insight in the episodes of analyzing and exploring a problem (Schoenfeld 1992; Veenman et al. 2005). Research has shown that schematic versus pictorial visual representations have good predictive validity for students' problem solving performance (Cox 1999; Edens and Potter 2007; Hegarty and Kozhevnikov 1999; Van Essen and Hamaker 1990; Van Garderen and Montague 2003). The correlation between the use of schematic visualizations and problem solving in mathematics ranges from about $r=0.3$ (explained variance 9 %) (Edens and Potter 2007) to about $r=0.7$ (explained variance 49 %) (Van Garderen and Montague 2003). So the predictive validity - the relation with problem solving performance as would be expected based on theory - of using the quality of problem visualizations as an indicator of metacognitive regulation seems to be in order. But, using problem visualizations as a metacognitive measure does not cover metacognition over all episodes of problem solving. To avoid underrepresentation of the construct, it is wise to add additional on-line information.

Another way to collect information on metacognitive processes on-line of the learning task is through performance (or calibration) judgments (Schraw 2009). More specifically, by assessing the accuracy of students' judgments of their own performance. The ability to judge one's performance has been conceptualized as an expression of metacognitive monitoring behavior (Boekaerts and Rozendaal 2010; Efklides 2006). When making on-line prediction judgments, that is to say estimations about performance before solving a problem, a student is especially concerned with the question whether he/she can analyze and categorize a problem. This gives the student a general idea whether he/she will be able to solve the problem or not. And a student may already briefly think ahead about a possible solution plan. There are also 'postdiction judgments' made after problem solving. By making a postdiction the student monitors if he/she has solved the problem correctly and adequately (Desoete 2009). Research has shown the accuracy of performance judgments before and after problem solving to have good predictive validity for mathematics performance. In the literature, correlations between judgments of performance and mathematics performance range from about $r=0.4$ to 0.6 (explained variance 16 % to 36 %) (Chen 2002; Desoete et al. 2001; Desoete 2009; Vermeer et al. 2000). The relationship is typically stronger when the performance measure is more closely related to the task on which the judgment is based (Pajares and Miller 1995). But, since accuracy measures give insight into a limited part of

metacognitive processes (monitoring by looking forward or looking backward and thinking ahead about a solution plan), it is recommendable to combine them with more measures of metacognitive regulation (Pieschl 2009), such as the type of visualizations students make.

What do we know about the overlap between different measures of metacognition? Sperling and colleagues (2004) compared the accuracy of performance judgments to the MAI self-report questionnaire. Their findings with college students show correlations around zero or even negative correlations between the accuracy of the performance judgments and the questionnaire. In the same vein, Veenman (2005) reviewed different studies and concluded that there is hardly any correspondence between findings from different on-line measures and self-report questionnaires. This shows that self-report instruments are generally not linked to students' on-line use of metacognition. On the other hand, we have little knowledge about the convergence between on-line performance judgments, problem visualizations and think-aloud scores. Theoretically, we can make some comparisons. As argued above, we expect the quality of problem visualizations to be specifically indicative for the way students analyze and explore a problem towards a solution plan. Such activities are indicators of metacognitive regulation in the first episodes of the problem solving process. Making performance judgments on the other hand primarily draws on students' metacognitive monitoring behavior and possibly on an initial stage of planning a solution. In think-aloud protocols, students' metacognitive regulation and monitoring are recorded over all episodes of problem solving. We would expect a low to moderate correspondence between performance judgments and think-alouds, since monitoring behavior is only a small part of all metacognitive processes executed when solving a problem (compare the findings on off-line performance judgments of Desoete 2008). And problem visualizations are not expected to cover metacognitive monitoring and regulation in the episodes of setting up and implementing a plan and verifying the solution, which are addressed in think-aloud protocols. So, theoretically, a think-aloud measure in word problem solving should show some overlap with visualizations, but should also have some unique predictive validity because it includes additional information about other problem solving episodes. Some additional differences between performance judgments and think-aloud scores may be caused by the fact that in think-alouds, metacognitive activities are measured which students perform without a specific assignment, while in the other on-line measures information is gathered about the quality of students metacognitive processes when instructed to perform certain actions. When comparing these different types of measures, it is important to use word problems with an adequate level of difficulty so students are enticed to use a varied set of metacognitive activities (Prins et al. 2006).

Since judgments of performance and problem visualizations theoretically measure different aspects of metacognition, but are both practical on-line measurement instruments with sufficient predictive validity, we suggest combining these measures into a new instrument. Collecting a combined measure of prediction judgments, postdiction judgments and visualizations of the problem on-line is meant to provide an indication of the intertwined process of metacognitive monitoring and regulation during problem solving. To study the relation of this newly combined measurement instrument with the other instruments discussed above we have formulated the following research questions:

- 1) What is the convergence between an on-line prediction-visualization-postdiction instrument, a self-report questionnaire and an on-line think-aloud instrument measuring metacognition?
- 2) Can the on-line prediction-visualization-postdiction instrument predict problem solving on an independent mathematical word problem test just as well as a think-aloud measure?

Based on the theoretical framework, we hypothesize there to be little to no convergence between the off-line, general self-report questionnaire and both on-line measures of metacognition in word problem solving. On the other hand, since the new instrument is collected as a practical on-line instrument measuring monitoring and regulation, we expect this instrument to have show moderate convergence with the on-line think-aloud measurement. But, because of the rich information in the think-aloud protocols, this measure is hypothesized to explain the largest proportion of variance in mathematical problem solving.

Method

Sample

The study reports of a total of 42 students randomly selected from five grade 5 classes in middle sized elementary schools. These students were in the business as usual condition of a larger study. We determined that the sample size is sufficient for detecting moderate correlations (between 0.30 and 0.40) (Cohen 1977). The average age of the students was 10.91 years old ($SD=0.28$). The sample consists of 24 boys and 18 girls. All students were of families with intermediate social economic status. Students scored a mean of 44.82 ($SD=5.61$) on the Raven Standard Progressive Matrices test, showing them to be well comparable to the norm scores in the Netherlands of 42 for the fiftieth percentile 47 for the seventy-fifth percentile (Raven et al. 1996). Over the days of testing, three students did not complete all measurements. So the effective sample is 39 students (22 boys, 17 girls).

Instruments

Think-aloud measure To collect think-aloud protocols, we used a ‘type 2’ procedure for verbal protocols (Ericsson and Simon 1993). This means we asked students beforehand to think aloud during execution of the word problems. After students started working on the problems, test leaders only interfered with neutral comments urging students to keep verbalizing (“keep thinking aloud”) when students silenced. Test leaders did not help the students to solve the problems in any way. The verbalizations of individual students’ thought processes were recorded using a video camera. This way, a detailed report of the verbalizations could be collected, without fully transcribing the protocols. The think-aloud data were gathered as follows.

First each student performed one test problem while thinking aloud. This was intended to help students get used to the procedure and the camera. This problem is not taken up in the analyses. During the actual measurement, students got two word problems (one by one) which they were instructed to solve while thinking aloud. Before starting, students got note paper and a pencil which they could use on their own initiative. Students were instructed in advance to indicate when they thought they were completely ready with the problem to make sure they were not stopped untimely by the test leader.

The two multistep problems used for the think-aloud protocols are presented below. Both problems lend themselves well for a metacognitive approach and they have multiple possible solution paths for reaching the correct answer. Moreover, both problems were judged by

three elementary school teachers as being rather difficult for fifth grade students so they specifically require a thoughtful approach (as opposed to atomized behavior).

<i>Hans and Ans are driving on the highway to Amsterdam.</i>	<i>Marie has bought a bag with 150 apples.</i>
<i>The highway has a gas station every 55 kilometers.</i>	<i>She wants to give all children of grade 5 as many apples as possible.</i>
<i>Their car breaks down after 196 kilometers.</i>	<i>Grade 5a has 13 children and in grade 5b there are 15 children.</i>
<i>Which gas station is the nearest, the previous one or the next?</i>	<i>Marie wants to give each child an equal amount of apples.</i>
	<i>She also wants to give 1 apple to the teacher of grade 5a and 1 apple to the teacher of grade 5b.</i>
	<i>How many apples will Marie have left?</i>

After having collected students' think-aloud protocols, each videotaped think-aloud session was assessed by two judges. The four judges received two hours of training in scoring the protocols. To rate the think-aloud protocols, a scoring scheme for systematical observation of think-aloud protocols was used (see Table 1). The scoring scheme was developed and tested by Veenman and colleagues (Veenman et al. 2000, 2005) and consists of activities which are characteristic for mathematical problem solving (Schoenfeld 1992). Previous research in secondary education has shown the instrument to be reliable and to have high convergent validity with full protocol analysis in which all verbalizations are transcribed.

Each activity in the scoring scheme was judged based on the verbal expressions of students while executing the word problems. Some verbalizations were thoughts preceding an activity, for instance when students verbalized how they were thinking about a plan before starting a calculation. Others thoughts were verbalized during the process, for

Table 1 Scoring scheme for systematical observation of think-aloud protocols in word problem solving

Episode	Activity
Read, analyze /explore (orientation)	1 Reading carefully
	2 Selection of relevant information/ numbers
	3 Paraphrasing the question
	4 Making a visualization or taking notes to orient on the task
	5 Estimating a possible outcome
Plan and implement (systematical orderliness)	6 Making a calculation plan
	7 Systematically executing the plan
	8 Being alert for correctness/ sloppiness (monitoring the calculation)
	9 Writing down calculations neatly
Verify (evaluation and reflection)	10 Monitoring the process
	11 Checking calculations and answers
	12 Drawing a conclusion
	13 Reflecting on the answer
	14 Reflecting on the learning experience

Items in bold print were used to compute a sumscore.

instance when students verbalized which information they selected from the text while doing so. Following the suggestion of the developers of the systematical observation scheme (Veenman et al. 2005), each activity was given a score ranging from 0 (not executed) to 1 (partially executed) to 2 (executed). An example of activity 6: Students got a score of 1 if they initiated a plan but do not follow through (for instance if a student would say “I am going to subtract” but gets distracted and does not carry on the planning into later solution steps). A score of 2 would be given for students who verbalize a worked out plan which they thought out before solving the problem (for instance saying: “First I need to subtract 13 by 5, and then I am going to divide by 2 to get the right answer”). Another example of activity 2: A student would get a score of 1 if he/she selects some numbers from the text but then quickly moves on (For instance by emphasizing information while reading aloud or by shortly repeating some of the numbers without concretely connecting this to a goal or plan). A score of 2 would be awarded if a student thoughtfully selects information for use in the calculation (For instance saying: “Let’s see, what do I need to calculate the answer? I need to know that every person gets 2 eggs and that there are 12 eggs in each box”).

The raters first watched the video of a word problem performed by a student (pausing and rewinding when needed) and individually filled in the scoring scheme. After this, they rewound the video and watched the problem solving of the student a second time, using the video data to explain each other which scores they gave and why. For each activity, the two raters argued until agreement was reached about the definitive scores before moving on to the next activity. This is a common approach in the scoring of think-aloud data (c.f. Elshout et al. 1993; Veenman et al. 2000, 2004). Observation of students’ scores on the items of the instrument shows that some regulation activities were not used by the relatively young students in the sample. For both word tasks, activities 3, 4, 5, 11, 13 and 14 showed little to no variance with almost all students scoring 0 points. These activities refer to sophisticated regulation processes such as reflection which are probably still underdeveloped for students in this early phase of development (Veenman et al. 2006). Leaving these items out leads to a maximum score of 16 points for the total instrument. Using the systematical observation scheme for the first twenty think-aloud protocols, a substantial interrater-reliability was found among the judges ($\kappa=0.95$, $p=0.00$).

VisA instrument As discussed in the theoretical framework, prediction judgments, postdiction judgments and problem visualizations were combined into one instrument. This instrument assesses a combination of metacognitive monitoring and regulation which are interrelatedly used during problem solving. We call this newly developed instrument the VisA instrument (Visualization and Accuracy). In the VisA instrument, four word problems are presented. For each word problem, students are asked to divide their problem solving over various steps:

- 1) Read the problem and rate your confidence for finding the correct answer (without calculating the answer);
- 2) Make a sketch which can help you solve the problem;
- 3) Solve the problem and fill in the answer;
- 4) Rate your confidence for having found the correct answer;

Four multistep word problems appropriate for using schematic visualizations were selected for the instrument. Students got approximately a maximum of five minutes to solve each problem. The four steps of each word problem are folded in the form of a booklet starting with step 1 as the front-page, step 2 and 3 on the middle two pages, and step 4 on the last page.

Figure 1 shows the first part of the instrument. Students are asked to fill in a traffic light with three options: Red (I am sure I cannot solve this problem), orange (I am not sure whether I will solve this problem correctly or incorrectly) and green (I am sure I will solve this problem correctly) and comment on the rationale for their answer. The latter is meant to have students think carefully and ask themselves why they think they can or cannot perform the task. Figure 1 also shows the second step of the instrument: Problem visualization. This step is presented on the inside of the booklet and was used to assess the quality of students' problem visualizations.

The scoring procedure for the instrument is designed to be straightforward so it is usable in research and practice. The scoring rules for each step are:

- 1) If students' prediction judgments are correct (i.e. students predicted they could solve the problem correct and indeed did; or they predicted they could not solve the problem and indeed gave the wrong answer) students get 1 point. If students' predictions are uncertain (orange traffic light) or incorrect (i.e. they predicted they could solve the problem correctly but in fact give the wrong answer; or they predicted they could not solve the problem but solved the problem correctly) they score 0 points.
- 2) For the visualization of the problem, students get 0 points if they made a pictorial sketch not depicting any of the important relationships in the problem, 0.5 point is awarded to sketches which are partly pictorial but have some schematic or mathematical features, and 1 point is given to primarily schematic visualizations.
- 3) The postdiction judgments of the students are scored in the same manner as step 1. Thus, students get 1 point when the postdiction is correct and 0 points when the postdiction does not match the answer.

After scoring all four word problems, a sum score was computed for the total instrument. The maximum score is 12 points. The first ten visualizations were scored with two judges

<p>Problem</p> <p>Marja plants rosebushes alongside the path to her house. The path is 27 meters long She plants a rosebush every 3 meters on both sides of the path. She also plants rosebushes at the beginning of the path.</p> <p>How many rosebushes does Marja need?</p> <p> Question</p> <p>How well do you think you can solve this problem?</p> <div data-bbox="150 1310 201 1460"> </div> <p> Please, explain why</p>	<p>Problem</p> <p>Marja plants rosebushes alongside the path to her house. The path is 27 meters long She plants a rosebush every 3 meters on both sides of the path. She also plants rosebushes at the beginning of the path.</p> <p>How many rosebushes does Marja need?</p> <p> Question</p> <p>Draw a sketch you can use to solve the problem.</p> <div data-bbox="612 1301 985 1543"> </div>
---	--

Fig. 1 Step 1 and 2 of the VisA instrument: Predicting one's performance and visualizing the problem situation

arguing until agreement about scoring rules for the visualizations was reached. Internal consistency of the instrument was $\alpha=0.70$.

Self-report questionnaire In this study, the ‘metacognitive self-regulation’ subscale of the MSLQ (Pintrich and De Groot 1990) is used. Statements in this subscale best match the metacognitive processes in the other instruments. This subscale contains 12 items in the form of statements about metacognitive behavior such as “Before I study new [mathematics] material thoroughly, I often read it through quickly to see how it is organized” And “When I execute [a math assignment], I set goals for myself in order to direct my activities.” General wording such as ‘in this course’ in the items were replaced by words specifically referring to mathematics.

Students were asked to indicate how much a statement applies to them by checking one out of five boxes ranging from ‘not at all true for me’ to ‘completely true for me’. Scores were coded ranging from no metacognitive regulation (not at all true for me: score 0) to a high amount of self-reported metacognitive regulation (completely true for me: score 4). Some items were stated in a reversed manner in the instrument but were recoded for the analyses. The maximum score on the instrument was 48 points. The internal consistency of the instrument was $\alpha=0.75$.

Mathematical word problem test As a performance measure, a test of 15 word problems was used. Of the test, two items with negative item-rest correlations were left out of the analyses. A sum score was calculated for the remaining 13 word problems. The test items are multistep word problems based on a national math assessment test (Janssen and Engelen 2002). Most students were familiar with the computations required to solve the problems. But, the fact that the computations are embedded in text turns them into word problems in which a metacognitive approach can benefit the solution process. Two examples of word problems from the test are presented below.

*Hassan already has € 250 in his savings account.
He is saving up for a game computer of € 490.
He saves € 40 each month
In how many months can Hassan buy the game
computer?*

*The pet store has a container with 5000 grams of dog
food.
Bart takes 30 % out for his dog.
How many grams of dog food stay in the container?*

Students got 1 point for each correct answer and 0 points for each incorrect answer. On average students in the sample solved 58 percent of the word problems ($SD=20$). The test had a reliability of $\alpha=0.65$.

Procedure

The word problem test and the self-report questionnaire were collected in the classroom with students filling in all questions individually. Subsequently, data were collected for the think-aloud measure and the VisA instrument. Half of the students completed the think-aloud measurement before the VisA measurement and the other half of the students completed the VisA before the think-aloud measure. Think-aloud protocols were collected individually in a quiet room outside of the classroom. Students completed the VisA measurement in a group setting.

Student responses that were missing after collecting the instruments (varying from 0.4 to 10.9 percent of the responses MCAR) were completed using the Expectation-Maximization Algorithm (Roth 1994; Schafer and Olsen 1998) in SPSS.

Results

Convergence between the instruments

In order to assess the convergence between the three measures aimed at measuring students' metacognition, means and bivariate correlations are presented in Table 2.

Students in our sample scored relatively low on all metacognitive measures, showing that metacognition is still in an early stage of development in upper elementary school. Concerning the relation with word problem solving, both on-line instruments – the think-aloud and the VisA instrument – were well related to performance with correlations ranging from $r = .57$ to $.48$. This is not the case for the self-report questionnaire which was not related to the mathematics test. Moreover, the self-report questionnaire showed no convergence with on-line metacognitive measures. Scores on the TA measure and the VisA instrument on the other hand were related, although the bivariate correlation is modest. Excluding one outlier with the highest TA score but a low VisA score would have led to a correlation between the two of $r(37) = .35$ and a correlation of VisA and PS of $r(37) = .50$ confirming that in general there is a moderate correlation between the two on-line instruments and that they are strongly related to problem solving performance.

Unique and shared predictive validity of think-aloud and VisA

To assess the amount of unique and shared explained variance of the think-aloud measure (TA) and the VisA instrument as predictors of scores on the word problem solving test, a regression commonality analysis was performed. Commonality analysis partitions a regression effect into unique and common effects. Unique effects show the amount of variance uniquely explained by a certain predictor variable. And common effects show how much explained variance two (or more) variables have in common (Nimon and Reio 2011). Results of the commonality analysis of the think-aloud measure and the VisA measure as predictors of problem solving performance are added in Table 3.

Table 3 shows in the first two columns that together the TA measure and the VisA measure correlated highly with problem solving performance ($r(37)=0.66$) and the variance explained by both measures was considerable (43 %). The data in columns three and four signify that TA and VisA have their own unique predictive value for performance. The beta coefficients indicate that 1 standard deviation change in TA score respectively VisA will lead

Table 2 Means and bivariate correlations between the different instruments measuring metacognition

	<i>M (SD)</i>	PS	TA	VisA
TA	9.87 (3.40)	0.57**	-	
VisA	4.15 (1.96)	0.48**	0.29*	-
SQ	25.52 (7.07)	0.03	0.16	-0.20

PS word problem solving test; TA Think-aloud measure; VisA VisA measure; SQ Student Questionnaire

* $p < 0.05$ ** $p < 0.01$

Table 3 Regression commonality analysis of a think-aloud measure and the VisA measure as predictors of word problem solving performance

Predictor	<i>R</i>	<i>R</i> ²	β	<i>p</i>	Unique	Common	Total
Both measures	0.66	0.43					
TA			0.48	0.00	0.21	0.12	0.33
VISA			0.34	0.01	0.11	0.12	0.23

TA Think-aloud measure; *VisA* VisA measure; *Unique* Unique variance explained by the predictor variable; *Common* Shared variance explained by both predictor variables; *Total* Total variance explained by the predictor

to a 0.48 respectively 0.34 standard deviation change in students' word problem solving. From the results in column five and six we can derive that both TA and VisA explained some unique variance in the word problem solving test (11 and 21 % respectively), and besides this they communally explained 13 percent of the variance in word problem solving. In total, TA explained 33 percent, and VisA 23 percent of the variance in word problem solving performance.

Conclusion and discussion

This study is intended as an exploration towards a more practical way of measuring metacognition to approximate the rich information of think-aloud protocols. Although imperative steps are being made towards new in-depth measures of metacognition supplementary to think-aloud protocol analysis (i.e. trace data; Azevedo et al. 2010; Greene and Azevedo 2010; Winne 2010), researchers and practitioners interested in students' metacognition still lack a practical instrument which is less complicated and time-consuming to use than think-aloud protocols or other in-depth measures (McNamara 2011). We suggest that one of the ways to make a step forwards in this issue, is evaluating findings on various instruments theoretically aimed at measuring metacognitive monitoring and regulation, and comparing their predictive and convergent validity (c.f. Veenman 2011a). Due to the fairly domain-specific nature of metacognition, we suggest the development of measurement instruments to be specifically molded to fit certain domains.

In this study, metacognition is measured in the domain of mathematical word problem solving. Findings from different measurement instruments were triangulated in an empirical study in grade five. Think-aloud observation was used as a comprehensive measure of metacognitive monitoring and regulation and as a reference point for other metacognitive measures. Think-alouds may not be appropriate for measuring automated processes (McNamara 2011; Veenman 2011b). But when collecting the protocols in an appropriate manner (Ericsson and Simon 1993), with tasks of a suitable level of complexity (Prins et al. 2006), think-alouds provide rich information on consciously used metacognitive processes. In our study, the think-aloud measure explains a total of 33 percent of the variance in mathematics performance, which is comparable to the predictive validity reported in other studies with think-aloud measures (Veenman et al. 2006). However, in the introduction we pinpointed the issue that collecting and analyzing think-aloud protocols is a very complex and time-consuming process. We reviewed several possible alternative measures which can be used in the design of a more practical, yet valid, measurement instrument. The empirical findings of our study using different measurement instruments are discussed below.

Firstly, based on the literature, it was hypothesized that a general off-line measures collected disconnected from the learning task would show little to no convergence with on-line measures. Findings of this study indeed support this claim. The self-report questionnaire we used shows no convergence with either the think-aloud measure nor the newly developed on-line instrument. Moreover, the questionnaire shows no relation to the problem solving test. This confirms the idea that what students say they do when asking them general self-report questions is not necessarily the same as what they actually do (Veenman 2011b). As discussed in the theoretical framework, this problem is likely to be caused by memory distortions as well as by variation in interpretation of the questions (McNamara 2011; Veenman 2011b). It could be that such issues can be addressed by fitting the formulation of the items more closely to the learning task (Schellings 2011). However, in our study we found that fitting the formulation of the questions to the learning domain (in this case by adding the word mathematics) does not seem to make the statements specific enough. Until we have more knowledge about how to increase concurrent and predictive validity of self-report questionnaires, we argue them to be more suitable as measures of metacognitive knowledge instead of on-line metacognitive metacognition which would be expected to directly influence performance (c.f. Desoete 2007; Greene and Azevedo 2010; Veenman 2005).

Secondly, we suggested to combine prediction judgments, problem visualizations, and postdiction judgments into a new instrument; the VisA instrument. All of these measures were argued to be indicative of metacognition, as well as having predictive validity for students' word problem solving performance. A large practical benefit of the VisA instrument is that it can be collected in paper and pencil format with groups of students. Teachers or test leaders need to make sure that students fill in every part of the instrument and do not inattentively skip parts. Another practical benefit of the instrument is that the scoring rules are quite straightforward to understand and use. How does the new instrument concur with scores collected with a think-aloud measure? Correlations between the new VisA instrument and think-aloud scores were predicted to be moderate since they both measure on-line metacognition, but the VisA instrument does not cover the whole range of metacognitive activities of the problem solving process captured in the think-alouds. Indeed, a moderate but significant relationship between the two on-line measures was found. The amount of metacognitive activities found with both on-line instruments is relatively low in our elementary sample. But, both instruments are significantly interrelated and are related to word problem solving performance.

Partialing out both instruments unique contribution as predictors of students' word problem solving in a regression commonality analysis, shows a substantial amount of shared predictive variance between the think-aloud measure and the VisA instrument. The overlap between the two instruments accounts for almost thirty percent of the total variance explained by both measures as predictors of word problem solving performance. This shows that in combining judgments of performance and problem visualizations, we have made a reasonable step forwards towards finding a valid and efficient instrument which corresponds to the think-aloud measure. Moreover, both instruments have predictive validity for word problem solving performance. As predicted, the think-aloud measure has the greatest predictive validity explaining 33 percent of the variance in problem solving performance. But, the VisA instrument also explains a sound part of 23 percent of the variance in word problem solving. The predictive validity of the VisA instrument for performance is comparable to the correlations reported previous studies of prediction and postdiction judgments (Chen 2002; Desoete et al. 2001; Desoete 2009; Vermeer et al. 2000) and problem visualizations (Edens and Potter 2007; Van Garderen and Montague 2003). The fact that the VisA instrument also uniquely covers some variance in problem solving which is not covered by

the think-aloud measure may be due to the fact that it measures metacognitive monitoring more strongly than the think-aloud measure in which monitoring is only represented in two of the sub-items (see “[Procedure](#)”). Also, the activities of drawing a sketch and making a prediction are not one-to-one related to the activities which students performed in the think aloud measure.

Concluding, our data confirms that think-aloud data gathered on-line of the problem solving process can provide much information about metacognitive processes affecting word problem solving. In searching for a more practical instrument, we found that the VisA instrument shows potential as an instrument for measuring metacognition in mathematical problem solving. The instrument has several benefits which facilitate data collection and scoring. Our empirical study has shown the VisA instrument to have predictive validity for mathematical word problem solving in elementary education. Additionally, the convergence with the think-aloud measure indicates that the instruments partly measure comparable constructs. However, the fact that VisA only partially overlaps with the think-aloud measure is a drawback. Depending on the breadth of the metacognitive construct one aims to measure, there may be more work needed to complete the puzzle.

One of the possible extensions of the present study is to further assess the convergent and predictive validity of performance judgments and visualizations by collecting them separately. Although there is already evidence for the predictive validity of separate prediction judgments and problem visualizations for performance, little is known about the convergence between these measures and other on-line measures such as think-aloud protocols. In VisA, substeps of the instrument are presented as interdependent steps of the problem solving process and can thus not be reliably detangled. But, in a follow-up study, it might be interesting to collect and compare independent measures of performance judgments and problem visualizations. Secondly, the use of think-aloud methods could be strengthened by using factor analysis to determine adequate scoring categories for the specific age group. Moreover, it would be interesting to add additional measures in a follow-up study to control for other variables possibly influencing findings of the think-aloud measure (i.e. verbal abilities; Veenman 2005) and the VisA instrument (i.e. spatial abilities; Cox 1999) in order to get a clearer picture of the constructs which are measured. This way, the theory about similarities and dissimilarities of different measures can be expanded. This can facilitate the search for new applied measures.

Although this first exploration of more practical measurement of metacognition in elementary education provides us with ground for further exploration, certain limitations must be kept in mind. Firstly, the measures in our design consist of quite few word problems. It would be well-advised to lengthen the measurement instruments with more word problems to increase their reliability. For instance, to increase the internal consistency of the VisA instrument up to $\alpha=0.80$, researchers might consider adding two or three comparable word tasks (Spearman 1910). Another more general limitation of most on-line measurement instruments is their obtrusive nature which might bias students’ responses in a certain direction (Schraw 2010). The amount of bias caused by the different obtrusive measures is not clear and should be taken into account when interpreting findings from think-aloud protocols and the VisA instrument.

Irrespective of the fact that there are clearly still some hurdles to be taken, we hope this study on more practical measurement of metacognition in word problem solving provides to be an incentive towards the exploration of more efficient, yet valid, measurement instruments in metacognition research. We believe this not only to be a valuable issue for researchers, but also for the community of practitioners interested in stimulating students’ metacognitive processes. Especially in schools where teachers have little time for testing individual students, it would be most relevant to have an efficient and valid instrument that

shows which teachable metacognitive skills some students lack and others have already acquired. Regarding to the large progress which has already been made in metacognitive theory development in the past decades, making the transition towards more practical use of our knowledge is an imperative - and exiting - step to take.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- 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 verbalisation method affect learning? *Metacognition and Learning*, 3, 39–58.
- Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction*, 20, 372–382.
- Brown, A. L., & DeLoache, J. S. (1978). Skills, plans, and self-regulation. In R. S. Siegler & R. S. Siegler (Eds.), *Children's thinking: What develops?* (pp. 3–35). Hillsdale: Lawrence Erlbaum Associates, Inc.
- Chen, P. P. (2002). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences*, 14, 77–90.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.
- Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, 9, 343–363.
- Cromley, J., & Azevedo, R. (2011). Measuring strategy use in context with multiple-choice items. *Metacognition and Learning*, 6, 155–177.
- Desoete, A. (2007). Evaluating and improving the mathematics teaching-learning process through metacognition. *Electronic Journal of Research in Educational Psychology*, 5, 705–730.
- Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: how you test is what you get. *Metacognition and Learning*, 3, 189–206.
- Desoete, A. (2009). Metacognitive prediction and evaluation skills and mathematical learning in third-grade students. *Educational Research and Evaluation*, 15, 435–446.
- Desoete, A., & Veenman, M. V. J. (2006). Metacognition in mathematics: Critical issues on nature, theory, assessment and treatment. In A. Desoete & M. V. J. Veenman (Eds.), *Metacognition in mathematics education* (pp. 1–10). New York: Nova Science Publishers, Inc.
- Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities*, 34, 435.
- Dignath, C., & Buttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, 3, 231–264.
- Edens, K., & Potter, E. (2007). The relationship of drawing and mathematical problem solving: "draw for math" tasks. *Studies in Art Education: A Journal of Issues and Research in Art Education*, 48, 282–298.
- Eklides, A. (2006). Metacognition and affect: what can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1, 3–14.
- Elshout, J. J., Veenman, M. V. J., & Van Hell, J. 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: Verbal reports as data* (rev. ed.). Cambridge: The MIT Press.
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, 137, 316–344.

- Fuchs, L. S., Zumeta, R. O., Schumacher, R. F., Powell, S. R., Seethaler, P. M., Hamlett, C. L., & Fuchs, D. (2010). The effects of schema-broadening instruction on second graders' word-problem performance and their ability to represent word problems with algebraic equations: a randomized control study. *The Elementary School Journal*, 110, 440–463.
- 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.
- Harskamp, E., & Suhre, C. (2007). Schoenveld's problem solving theory in a student controlled learning environment. *Computers in Education*, 49, 822–839.
- Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology*, 91, 684–689.
- Jacobse, A. E., & Harskamp, E. G. (2009). Student-controlled metacognitive training for solving word problems in primary school mathematics. *Educational Research and Evaluation*, 15, 447–463.
- Janssen, J., & Engelen, R. (2002). *Verantwoording van de toetsen rekenen-wiskunde 2002 [Account of the mathematics tests 2002]*. Arnhem: Citogroep.
- Kramarski, B., & Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments? *Journal of Computer Assisted Learning*, 22, 24–33.
- McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: challenges and successes. *Metacognition and Learning*, 6, 195–203.
- Mevarech, Z. R., & Amrany, C. (2008). Immediate and delayed effects of meta-cognitive instruction on regulation of cognition and mathematics achievement. *Metacognition and Learning*, 3, 147–157.
- Muis, K. R., Winne, P. H., & Jamieson-Noel, D. (2007). Using a multitrait-multimethod analysis to examine conceptual similarities of three self-regulated learning inventories. *The British Journal of Educational Psychology*, 77, 177–195.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, 51, 102–116.
- Nimon, K., & Reio, T. G., Jr. (2011). Regression commonality analysis: a technique for quantitative theory building. *Human Resource Development Review*, 10, 329–340.
- Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: the need for specificity of assessment. *Journal of Counseling Psychology*, 42, 190–198.
- Pieschl, S. (2009). Metacognitive calibration - an extended conceptualization and potential applications. *Metacognition and Learning*, 4, 3–31.
- 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.
- Prins, F. J., Veenman, M. V. J., & Elshout, J. J. (2006). The impact of intellectual ability and metacognition on learning: new support for the threshold of problematicity theory. *Learning and Instruction*, 16, 374.
- Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for raven's progressive matrices and vocabulary scales. Section 3: Standard progressive matrices*. Oxford: Oxford Psychologist Press.
- Roth, P. L. (1994). Missing data: a conceptual review for applied psychologists. *Personnel Psychology*, 47, 537–560.
- Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behavioral Research*, 33, 545–571.
- Schellings, G. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning*, 6, 91–109.
- Schellings, G., & Van Hout-Wolters, B. H. A. M. (2011). Measuring strategy use with self-report instruments: theoretical and empirical considerations. *Metacognition and Learning*, 6, 83–90.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching* (pp. 224–270). New York: McMillan Publishing.
- Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4, 33–45.
- Schraw, G. (2010). Measuring self-regulation in computer-based learning environments. *Educational Psychologist*, 45, 258–266.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19, 460–475.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271–295.
- 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. C., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation*, 10, 117–139.
- Van der Stel, M., & Veenman, M. V. J. (2010). Development of metacognitive skillfulness: a longitudinal study. *Learning and Individual Differences*, 20, 220–224.

- Van Essen, G., & Hamaker, C. (1990). Using self-generated drawings to solve arithmetic word problems. *The Journal of Educational Research*, 83, 301–312.
- Van Garderen, D., & Montague, M. (2003). Visual-spatial representation, mathematical problem solving, and students of varying abilities. *Learning Disabilities Research and Practice*, 18, 246–254.
- 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. 77–99). Münster: Waxmann.
- Veenman, M. V. J. (2011a). Alternative assessment of strategy use with self-report instruments: a discussion. *Metacognition and Learning*, 6, 205–211.
- Veenman, M. V. J. (2011b). Learning to self-monitor and self-regulate. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197–218). New York: Routledge.
- Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: age and task differences. *Learning and Individual Differences*, 15, 159–176.
- Veenman, M. V. J., & Van Hout-Wolters, B. (2002). Het meten van metacognitieve vaardigheden [Measuring metacognitive skills]. In F. Daems, R. Rymenans, & G. Rogiest (Eds.), *Onderwijsonderzoek in Nederland en Vlaanderen. Proceedings 29e ORD 2002* (pp. 102–103). Wilrijk: Universiteit van Antwerpen.
- Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: availability versus production deficiencies. *Anxiety, Stress and Coping*, 13, 391.
- 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.
- Veenman, M. V. J., Kok, R., & Blöte, A. W. (2005). The relation between intellectual and meta-cognitive skills in early adolescence. *Instructional Science: An International Journal of Learning and Cognition*, 33, 193.
- 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.
- Vermeer, H. J., Boekaerts, M., & Seegers, G. (2000). Motivational and gender differences: sixth-grade students' mathematical problem-solving behavior. *Journal of Educational Psychology*, 92, 308–315.
- Weinstein, C. E., Zimmermann, S. A., & Palmer, D. R. (1988). Assessing learning strategies: The design and development of the LASSI. In C. E. Weinstein, E. T. Goetz, P. A. Alexander, C. E. Weinstein, E. T. Goetz, & P. A. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction, and evaluation* (pp. 25–40). San Diego: Academic.
- Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, 45, 267–276.
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, M. Zeidner, M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego: Academic.