Measuring strategy use with self-report instruments: theoretical and empirical considerations

Gonny Schellings · Bernadette Van Hout-Wolters

For many years educational research and practice have paid attention to learners' learning strategies and their improvement. Both researchers and practitioners use different kinds of measuring methods to assess learning strategies (Van Hout-Wolters 2000, 2009; Veenman 2011; Veenman et al. 2006; Winne and Perry 2000). Learning strategies can be defined as certain combinations of goal-oriented activities applied to learning in school settings. A distinction is regularly made between cognitive, metacognitive and affective learning strategies. Numerous specific learning strategies are distinguished, and diverse overviews of learning strategies have appeared in several publications (see e.g., Alexander 2006; Pintrich 2004; Van Hout-Wolters et al. 2000; Winne and Hadwin 1998).The ongoing focus on conceptualizing learning strategies, especially metacognitive strategies, is paralleled by ensuing debates about how to measure these strategies, and the usability of self-report instruments is particularly prone to heated scientific discussions. This special issue is contributing to the discussion of strategy assessment by presenting five empirical studies in which pros and cons of self-report instruments are demonstrated. All articles attend to validity issues surrounding the measurement of cognitive and metacognitive strategies with self-report instruments.

The reasons for measuring learning strategies are diverse. For example, teachers may administer questionnaires to obtain information about the strong and weak points of learners' strategies and provide the learners with individual support accordingly. Learning strategies can also be measured to determine whether the formulated (sub)goals of strategy

G. Schellings (✉) · B. Van Hout-Wolters
Research Institute of Child Development and Education, University of Amsterdam,
Spinozastraat 55, 1018 HJ Amsterdam, The Netherlands
e-mail: g.l.m.schellings@uva.nl

B. Van Hout-Wolters
e-mail: b.h.a.m.vanhout-wolters@uva.nl

instruction have been reached (see e.g., Mevarech and Fridkin 2006; Aarnoutse and Schellings 2003). Furthermore, measuring methods are used in non-evaluative ways; for example, in exploring learning strategies that are executed during studying expository texts (e.g., Schellings et al. 2006), or to examine whether specific strategies are related to the learning results in a certain type of learning task (e.g., Broekkamp and Van Hout-Wolters 2007; Hadwin et al. 2001; Van der Stel and Veenman 2008).These different goals for measuring learning strategies can render different conditions (i.e. methodological or practical) to which measuring methods must adhere. Before introducing the research papers in this special issue, we present some issues surrounding the decision to administer self-report instruments in educational practice; next we discuss methodological considerations in self-report instrument research.

Administering self-report instruments: considerations in selection

Self-report and Likert-type instruments constitute regular methods of measuring learning strategies (e.g. Van Hout-Wolters 2009; Veenman 2011). The most common characteristic of self-reports is that learners themselves classify or deduce their own activities. The advantages of self-reports, for example questionnaires, are obvious: the learners are not disturbed during their learning activities and questionnaires are easy to administer in large-scale testing. The disadvantages, however, may be that learners are not able to accurately recollect what they have done (Veenman 2005).

Notwithstanding the methodological drawbacks of self-report instruments, these instruments are widely used without in depth consideration of the nature and quality of these measures to uncover individuals' learning behavior. Well-known questionnaires do differ in content (cf. Muis et al. 2007). In administering a self-report instrument, it is important not to simply select a popular one (e.g., MSLQ, Pintrich and De Groot 1990; MAI, Schraw and Dennison 1994), but to question exactly which learning strategies have to be measured and which learning strategies are measured by the chosen instrument (cf. Van Hout-Wolters 2009). Furthermore, it is important for the selection of an appropriate self-report instrument to know at which specific learning task it is aimed. And, additionally important, that this specific learning task is representative of the tasks that will be generalised.

After defining the aims of the measurement, i.e. which learning strategies are to be measured for which learning tasks, the use of a self-report instrument faces other methodological requirements, such as the standard requirements of validity and reliability (cf. Cohen et al. 2007; Winne et al. 2002). Additionally, the issue of the generalisability of measuring methods is important. Winne and Perry's (2000) point of departure is that 'every measurement is a sample of behavior'. All answers to questionnaires and interviews can be regarded as 'selections of information from memory' and much remains unrevealed. For example, which learning situation does a learner have in mind when he is completing a general learning strategy questionnaire? How does the learner perceive the importance or the frequency of his learning strategy? However, the issue of generalisability is present for all kinds of measuring methods rather than an exclusive issue for self-report questionnaires. For example, if a learner in a think-aloud session does not mention a certain learning activity (e.g. integrating information) the questions of whether he is unable to do this, whether he can do this but does not think of doing so, or whether he can do it but decides not to do it here remain. A methodological point of discussion, therefore, is that too little is known about how measuring methods and self-report instruments in particular lead to responses and how we should interpret them (Van Hout-Wolters 2009).

Apart from the methodological considerations, self-report instruments are often used because of their practical usefulness. Particularly when teachers or tutors are selecting a measuring method for their own school practice, they usually choose Likert-type scales. But also in research, practical considerations lead to administering self-report measures; for example, the size of the sample, and the age and ability of learners. When group measurement is the point of departure, then often not oral, but written or digital self-reports are gathered because these instruments take much less time in entire classes or larger groups than individual measuring methods. Moreover, the processing of the self-report data (especially questionnaires) is also less labor-intensive and cost-effective.

Because self-report instruments have their strong and their flawed points, many researchers recommend using a combination of measuring methods. For example, Dinsmore et al. (2008) conclude about the "measurement conundrum" found in their review of 255 studies that neither quantitative (e.g., self-report surveys) nor qualitative approaches (e.g., think-aloud studies) will suffice to reveal the self-regulated learning strategies, but that some combination may be required. With this *multi-method use*, also known as mixed methods or triangulation (Burke Johnson and Onwuegbuzie 2004; Winne and Perry 2000), the power of each method is taken to obtain a broad picture and deep insight into learners' learning strategies. However, different types of measuring methods may lead to a variety of results and the question of how to judge them arises. Do some results have greater value than others and could these results be combined? Additionally, using a multi-method design and interpreting the combination of data may become a rather labor-intensive and cost-consuming undertaking.

Examining self-report instruments: methodological considerations

The previous paragraphs dealt with choosing an appropriate instrument for measuring learning strategies (especially self-report instruments or questionnaires), but the measurement itself is increasingly becoming the topic of research and scientific publications. Greene and Azevedo (2010) mention the appearance of a number of special issues of educational or learning science journals that are dedicated to the discussion of how self-regulated learning processing, in general, should be conceptualized and the implications of those conceptualizations for measurement. Also at several conferences many papers are presented and symposia are organized around this topic of methods for measuring cognitive and metacognitive processing in learning settings. For example, at the biennial EARLI-conference 2009 in Amsterdam, there were two symposia organized about measuring (meta)cognitive learning strategies and multi-method approaches in assessing self-regulated learning. The validity of a questionnaire's data is especially challenged: learners may not be able to accurately report what they generally do or what they have done in finishing an assignment; or questionnaires may measure the learners' perception rather than strategies actually performed (cf. Richardson 2004; Perry and Winne 2006). Veenman (2005, 2011) argues that the utility of off-line methods such as self-report instruments should be reconsidered and, for the time being, on-line methods should be preferred. Off-line methods show low concurrent validity with on-line measures and they do not predict learning outcomes very well. Accordingly, Veenman assumes that self-reports do not measure actual activities performed, rather that they might assess knowledge of those activities. Furthermore, such knowledge does not need to imply that learners will actually apply those activities.

Notwithstanding this criticism, we believe that the possibilities of large-scale testing and the practical usefulness should not be underestimated. Each method may have its own

quality, but further research is needed to develop more precise measures. This type of research may not only be aimed at comparing different methods in one research design (i.e., multi-method research), but also at the characteristics of self-report instruments in order to improve the instruments or to construct alternative assessment measures with the advantages of self-reports.

In *multi-method research* (cf. e.g., Veenman 2005; Winne and Perry 2000) the goal is to gain more insight into the strengths and weaknesses of the individual methods that are administered within one research design. That is to say, the instruments are provided to the same participants in the same learning situation. In comparing different instruments, the methodological issues of interest concern convergent validity, discriminant validity, method bias in the instrument itself, significant method variance, and measures of calibration. Besides, multi-method research should not be limited to the analyses of total scores of measuring methods and to correlation analyses. High correlations could, after all, be ascribed to background factors. The measuring methods to be compared should also entail the measurement of the same specific learning strategies in a similar learning situation. So this type of research should be deepened; that is to say, more sophisticated designs and analyses in comparing the instruments should be taken into account. On the basis of this type of multi-method research self-report instruments could be improved. And it will also render more theoretical insight into the field of metacognition and self-regulated learning behavior (cf. Dinsmore, et al. 2008; Veenman, et al. 2006; Winne and Perry 2000).

Another line of research to improve the self-report instruments is examining the measuring characteristics of the self-reports. In the first place, one must have a clear idea of what the self-report instruments are measuring, since they differ in content as Muis and her colleagues (2007) found in their study. Four facets of self-regulated learning they conceptually identified as common constructs across three popular questionnaires appeared to be empirically disparate. At the same time, researchers question the preference of measuring learning activities during the learner's learning (on-line) or apart from it (off-line, that is, when the learner is not learning) as in self-report instruments. The key difference in measurement may not be the "real task-involvement" but the "task-specific reporting". Most self-report instruments, especially the questionnaires, concern very broad-brush measures that seek to generalize across multiple times and situations, as well as across cognitive, motivational, emotional and behavioral domains (cf. Muis, et al. 2007; Richardson 2004). Accordingly, Dinsmore et al. (2008) wonder how these instruments can fairly and accurately measure learning strategies or capture the dynamic interplay of person, environment, and behavior that is the hallmark of self-regulated learning. Task-specific measuring connects to ideas and research, from which it appears that learners' learning strategies differ for types of learning tasks or subjects (cf. Samuelstuen and Bråten 2007; Broekkamp and Van Hout-Wolters 2007; Hadwin et al. 2001). However, the distinction in measuring within or outside the context (i.e., task-specific or general measures) indirectly relates to the distinction in on-line versus off-line measuring. On-line measuring is, by definition, bound to the task performed within the assessment. Off-line measures can be aimed at general learning or learning from one specific task. In order to deepen the discussion of validity issues surrounding self-reports, such as questionnaires, instruments should be tailored to particular tasks or contexts (cf. Samuelstuen & Braten, 2007; Richardson 2004). In addition, the precise moment of collecting self-reports may be examined; for example, is it possible to gather self-reports straight after or even during task execution.

With the advancements in technology and with the acceptance of multi-method studies, some alternative assessment measures may be examined. For example, Van Gog et al. (2005) use eye-tracking to cue retrospective self-reports. The learners are asked what they

thought during task execution while they watch the recordings of their eye-movements and mouse-keyboard actions. Karabenick et al. (2007) use cognitive interviewing; that is, systematically interviewing the respondent after answering each individual item of a questionnaire. Magliano and colleagues (this issue) construct a computer tool that deduces reading strategies. In using the tool, the learners are asked self-report questions and the tool digitally labels the strategies mentioned.

In sum, the discussion about using self-reports to measure cognitive and metacognitive strategies is vivid; and more theoretical and practical considerations may be taken into account, as has been done in the studies that are included in this special issue.

Overview of articles in this special issue

This special issue brings together papers presented at the EARLI 2009 symposium "Measuring learning strategies: What are we measuring?" All contributions are searching for some kind of validation data concerning measuring learning strategies with self-report instruments or instruments derived from self-report measures. The five articles are followed by two commentaries by Danielle McNamara and by Marcel Veenman. In her commentary, McNamara also discusses some educational implications. Veenman especially comments on important methodological concerns.

In the first contribution a multi-method design is used. Schellings compares a task-specific questionnaire with the think-aloud method. The aim of this study is to examine whether a more sophisticated design will result in a higher correlation between the questionnaire and think-aloud protocols than regularly reported. Both the questionnaire and the think-aloud method are directed at the same learning activities in the same learning task (i.e., studying a history text). In particular, the questionnaire is directly based on a taxonomy for coding think-aloud protocols in text studying. The questionnaire results are correlated with think-aloud protocol results. Furthermore, a case study is performed in which students first study the text while thinking aloud and then continue to think aloud while completing the questionnaire. As a result, differences between the ratings of activities on both instruments were compared in more depth.

In the second article by Bråten and Strømsø, a task-specific questionnaire is also used. This questionnaire is directed at readers' strategies exerted during comprehending and making connections across multiple expository texts (multiple-text paradigm). In their study, the researchers have constructed a self-report inventory focusing on strategic multiple-text processing in that specific task context (reading separate texts on a science topic). Respondents are instructed to monitor the strategies that they use while reading the texts and they are told they would receive some questions about what they did during reading. Through factor analysis the inventory reveals two dimensions of multiple-text comprehension strategies: one concerns the accumulation of pieces of information from the different texts; the other concerns elaboration across the different texts. These dimensions are compared with intratextual and intertextual comprehension measures, while taking the possible intervening role of prior knowledge into account.

Both Schellings and Bråten & Strømsø use a retrospective, off-line self-report instrument, whereas Magliano, and his colleagues Millis, the RSAT Development Team, Levinstein and Boonthum, examine an on-line instrument (third contribution). Their instrument is a computer-based reading assessment designed to measure readers' comprehension and spontaneous use of reading strategies while reading texts. In the tool, readers comprehend passages one sentence at a time, and are asked an indirect self-report

("What are your thoughts regarding your understanding of the sentence in the context of the passage?") or a direct question (e.g., "Why X?") after reading each pre-selected target sentence. The indirect question is used to elicit self-reports from the readers about their strategies; the direct questions are designed to provide an assessment of comprehension. The answers are digitally analyzed through counting target words that are associated with strategies or with comprehension. The measures resulting from the RSAT tool are used to predict comprehension scores on standardized tests.

An on-line instrument which focuses on strategy performance is examined in the fourth contribution presented by Cromley and Azevedo. In their search for an alternative for self-report questionnaires but maintaining their important advantages, i.e., group administration and simple scoring, they develop a domain-general and domain-specific multiple-choice strategy use measure. This method requires readers to enact a strategy. It measures whether readers perform the strategy accurately by posing multiple-choice questions. Cromley and Azevedo examine their instrument with three samples as part of a series of studies testing a model of reading comprehension. In the three studies, their instrument is correlated with standard reading comprehension and component measures of reading, including vocabulary, word reading, inference and background knowledge. In addition to searching for the convergent validity among the different instruments, commonality analysis is also performed to examine to what degree strategy use may contribute to single-text comprehension.

A more methodological query is presented in the last contribution of this special issue. Winne and Muis explore the validity of two measures of calibration (judgment accuracy). Learners' accuracy in estimating what they know is critical to learning effectively. In essence, accuracy ratings are a type of self-report used in conjunction with a specific task. Moreover, they not only contextualize learners' estimation of their cognitive ability, but they also involve a kind of test to correlate performance estimates with real outcomes. Winne and Muis provide their participants with three knowledge tests, after which the learners judge the correctness for each answer provided. Because of distributional assumptions that data challenge, researchers generally prefer the Goodman-Kruskall gamma coefficient (G) as the appropriate measure of calibration rather than the signal detection theory's d' statistic. Winne and Muis question this by reviewing literature and by empirically comparing G and d'. Furthermore, they examine whether a learner's calibration varies across three domains of knowledge.

In all, the five different contributions all examine theoretical, methodological and practical issues concerning measuring strategy use by self-report instruments or their derivates. The instruments differ in off-line versus on-line measuring, but all instruments that are described relate to task-specific measuring. After the EARLI-2009 symposium "Measuring learning strategies: What are we measuring?", the presented papers were extended and elaborated into contributions handling diverse kinds of validation data concerning self-report instruments for measuring learning strategies and issues of calibration. By bringing together these contributions, new insights may evolve concerning self-report instruments. All articles contribute to the search for practically useable learning strategy measures suited for large group administration.

References

Aarnoutse, C. A. J., & Schellings, G. L. M. (2003). Learning reading strategies by triggering reading motivation. *Educational Studies, 29*, 387–409.

Alexander, P. A. (2006). *Psychology in learning and instruction*. Upper Saddle River (NJ): Pearson Prentice Hall.

Broekkamp, H., & Van Hout-Wolters, B. H. A. M. (2007). Students adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review, 19*, 228–401.

Burke Johnson, R., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14–26.

Cohen, L., Manion, L., Morrison, K. (2007). *Research methods in education*. (6th ed.) Routledge Falmer

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.

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.

Hadwin, A. L., Winne, P. H., Stockley, D. B., Nesbit, J. C., & Woszczyna, C. (2001). Context moderates students' self-reports about how they study. *Journal of Educational Psychology, 93*, 477–487.

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., et al. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42*(3), 139–151.

Mevarech, Z., & Fridkin, S. (2006). The effects of IMPROVE on mathematical knowledge, mathematical reasoning, and metacognition. *Metacognition and Learning, 1*, 85–97.

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. *British Journal of Educational Psychology, 77*, 177–195.

Perry, N. E., & Winne, P. H. (2006). Learning from learning kits: gStudy traces of students' self-regulated engagements with computerized content. *Educational Psychology Review, 18*, 211–228.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385–408.

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.

Richardson, J. T. E. (2004). Methodological issues in questionnaire-based research on student learning in higher education. *Educational Psychology Review, 16*, 347–358.

Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *British Journal of Educational Psychology, 77*, 351–378.

Schellings, G. L. M., Aarnoutse, C. A. J., & Van Leeuwe, J. (2006). Third-grader's think aloud protocols: Types of reading activities in reading an expository text. *Learning and Instruction, 16*, 549–568.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460–475.

Van der Stel, M., & Veenman, M. J. V. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learning and Individual Differences, 18*, 128–134.

Van Gog, T., Paas, F., Van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology. Applied, 11*(4), 237–244.

Van Hout-Wolters, B. H. A. M. (2000). Assessing self-directed learning. In P. R. J. Simons, J. van der Linden, & T. Duffy (Eds.), *New learning* (pp. 83–101). Dordrecht: Kluwer.

Van Hout-Wolters, B. H. A. M. (2009). Leerstrategieën meten. Soorten meetmethoden en hun bruikbaarheid in onderwijs en onderzoek. [Measuring Learning strategies. Different kinds of assessment methods and their usefulness in education and research]. *Pedagogische Studiën, 86*, 103–110.

Van Hout-Wolters, B. H. A. M., Simons, P. R. J., & Volet, S. (2000). Active learning: Self-directed learning and independent work. In P. R. J. Simons, J. L. van der Linden, & T. M. Duffy (Eds.), *New learning* (pp. 21–37). Dordrecht: Kluwer.

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multimethod designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 75–97). Berlin: Waxmann.

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., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3–15.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale: Erlbaum.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In P. R. Pintrich, M. Boekaerts, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–567). Orlando: Academic.

Winne, P. H., Jamieson-Noel, D. L., & Muis, K. (2002). Methodological issues and advances in researching tactics, strategies, and self-regulated learning. In P. R. Pintrich & M. L. Maehr (Eds.), *Advances in motivation and achievement: New directions in measures and methods, vol. 12* (pp. 121–155). Greenwich: JAI.