




Granularity matters: comparing different ways of measuring self-regulated learning

Sanne F. E. Rovers¹  · Geraldine Clarebout¹ · Hans H. C. M. Savelberg¹ · Anique B. H. de Bruin¹ · Jeroen J. G. van Merriënboer¹

Received: 18 July 2017 / Accepted: 1 February 2019 / Published online: 12 February 2019
© The Author(s) 2019

Abstract

Although self-regulated learning (SRL) is becoming increasingly important in modern educational contexts, disagreements exist regarding its measurement. One particularly important issue is whether self-reports represent valid ways to measure this process. Several researchers have advocated the use of behavioral indicators of SRL instead. An outstanding research debate concerns the extent to which it is possible to compare behavioral measures of SRL to traditional ways of measuring SRL using self-report questionnaire data, and which of these methods provides the most valid and reliable indicator of SRL. The current review investigates this question. It was found that granularity is an important concept in the comparison of SRL measurements, influencing the degree to which students can accurately report on their use of SRL strategies. The results show that self-report questionnaires may give a relatively accurate insight into students' global level of self-regulation, giving them their own value in educational research and remediation. In contrast, when students are asked to report on specific SRL strategies, behavioral measures give a more accurate account. First and foremost, researchers and practitioners must have a clear idea about their research question or problem statement, before choosing or combining either form of measurement.

Keywords Self-regulated learning · Trace data · Self-report · Granularity · Calibration

Since the emergence of the Internet, there has been a sharp increase in the amount of information available, and the half-life of information has dramatically shortened (Arbesman 2013). For students and professionals alike, this “information jungle” can be hard to navigate. In order to survive, they need to constantly monitor and evaluate their progress towards their own learning goals, and adjust their behavior if necessary. These skills are captured by so-called *self-regulated learning skills* (Zimmerman and Schunk 1989). Within the context of

✉ Sanne F. E. Rovers
s.rovers@maastrichtuniversity.nl

¹ School of Health Professions Education, Maastricht University, Universiteitssingel 60, PO Box 616, 6200 MD Maastricht, The Netherlands

education, self-regulated learning (SRL) concerns the process whereby learners actively take charge of their own learning. They actively monitor their learning process and outcomes, and are able to regulate and adapt their behavior, cognition and motivation when necessary to optimize their learning outcomes (Zimmerman 2000). Due to the enormous increase in available information, SRL has become much more important for students in order to complete their education, but its measurement has been a problematic aspect (Winne 2010; Veenman et al. 2006).

In order for students to develop effective SRL strategies and to be appropriately supported in this development, researchers and educators need accurate, reliable measures for this construct. In this way, researchers and educators can come to an accurate account of students' self-regulation and students' points for improvement. In 2008, Zimmerman published an article on innovative ways of measuring SRL (Zimmerman 2008). In this work, several outstanding questions regarding the measurement of SRL were highlighted. One of these questions was the extent to which it is possible to compare trace data (computerized log files of students' online behaviors) to traditional ways of measuring SRL using self-report questionnaire data. As Veenman (2005) pointed out, convergence between self-report statements and concurrent behavior tends to be low, however many new approaches have emerged since this publication that warrant closer investigation. This narrative review will discuss the progress that has been made in this area and extend the question to the comparison of self-report questionnaires to several online measures of SRL.

To provide the background for this study, this section will first introduce some of the most influential models describing the SRL process. Ideally, measurement of SRL is informed by a theoretical model, with the model serving as the underlying framework. After introducing the models, we will describe some important considerations in the measurement of SRL (online versus offline measuring, calibration and granularity), after which we will describe the common methods used for measurement. Being the traditional way to measure SRL (Schellings and Van Hout-Wolters 2011), we will first introduce the use of self-report questionnaires. We will then address the concerns associated with this form of measurement, before describing alternative, online forms of measurement (e.g. think-aloud protocols, systematic observations and computerized log data).

Influential models of SRL have been put forth by Zimmerman (2000), Pintrich (2004) and Winne and Hadwin (1998). Social-cognitive models were developed by Zimmerman (2000) and Pintrich (2004). Another SRL model was developed by Winne and Hadwin (1998) and focuses on the specific cognitive processes that occur during learning, such as memory processes and operations (Greene and Azevedo 2007).

A common theme in most SRL models is that SRL is viewed as a loosely sequenced process of cyclical phases. The social-cognitive model postulated by Zimmerman (2000) describes a cyclical feedback loop of three phases constituting SRL: forethought, performance and reflection. The cyclical nature of these phases postulates that the outcome of each phase provides input for and influences processes in the other phases.

Pintrich (2004) put forth a social-cognitive model of SRL that posits motivation, self-efficacy and goal orientation as the discerning aspects of SRL. His model consists of four phases similar to those put forth by Zimmerman (2000), including forethought, monitoring, control and reflection. As an addition to Zimmerman's model, the model postulates four areas for regulation of learning. Specifically, students can regulate their cognition, motivation, behavior and learning environment.

The SRL model by Winne and Hadwin (1998) consists of four phases: task definition, setting learning goals and plans, enactment of learning strategies, and adapting. In each of these phases, SRL is influenced by a set of processes involving interaction between the conditions, operations, products, evaluations and standards (COPEs) that students find themselves in. The learning cycle within this model is again loosely sequenced: learners are expected to go through all of the phases, but may return to earlier phases when they feel this will help improve their products in a later phase. Throughout the process, learners apply a range of choices based on their motivation to execute the task at hand (Winne 2017).

When it comes to the measurement of SRL, several considerations are reflected in the different forms of measurement. First of all, a distinction can be made between online and offline measures (Schellings 2011; Veenman 2005), depending on the timing of the measurement. Online measures (sometimes called process measures) take place during the performance of the actual learning task. Examples include think-aloud protocols, systematic observations or computerized traces with log data. Offline measures are collected either before or after task performance. Self-report questionnaires usually fall into this category. It is important to note that, although this terminology can be somewhat misleading, the distinction between online and offline measurements does not refer to the mode of administration (i.e., whether or not the Internet is used), but to the timing of measurement (before/after or during task performance). For example, a questionnaire that is administered electronically *after* task performance will still be considered an offline measurement, while a micro-analytic question administered on paper *during* task performance is an online measurement.

Another important construct in the context of this review is the concept of calibration, which can be defined as the degree of correspondence between an individual's self-report of a certain cognitive construct versus the actual, online value of this construct (Winne and Jamieson-Noel 2002). This calibration can focus on process variables, for example on the degree of overlap between students' self-reports of cognitive strategy use and their actual strategy use, or on outcome variables, such as the level of correspondence between students' estimated achievement and their actual achievement (e.g. judgments of learning, Schneider 2008). The focus of this review is on calibration in terms of process variables.

Finally, SRL can be measured at different levels of granularity. Granularity refers to level of detail at which self-regulatory processes are measured. SRL can be measured on a coarse grained level when looking at global SRL process phases, as opposed to fine grained SRL measurements that focus on students' micro-level SRL processes (Azevedo 2009).

Reflecting these different considerations at varying levels, different methods have been applied to measure SRL. The traditional way to measure SRL is through self-report questionnaires. Examples include the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al. 1991), the Learning and Study Strategies Inventory (LASSI; Weinstein and Palmer 2002) and the Metacognitive Awareness Inventory (MAI; Schraw and Dennison 1994). Self-report questionnaires tend to treat SRL as a stable aptitude or trait belonging to an individual, giving an indication of how an individual *usually* approaches learning tasks, thereby aggregating these approaches to studying across studying contexts, episodes and tasks (Schellings 2011; Schellings et al. 2013; McCardle and Hadwin 2015). A main reason for the popularity of self-report questionnaires is the ease with which they can be administered and analyzed, making it possible to examine large samples of learners (Schellings and Van Hout-Wolters 2011). However, increasingly large numbers of researchers in the field have stated objections to this approach to measuring SRL (Veenman 2005; Winne et al. 2002; Winne and Perry 2000). These criticisms can be roughly divided into two concerns.

The first concern regards the treatment of SRL as a dynamic and context dependent process versus a static and stable trait. SRL is considered to be a context-dependent process, and SRL strategies employed by students may vary both across and within learning tasks and contexts (McCardle and Hadwin 2015; Winne and Hadwin 1998; Bråten and Samuelstuen 2007). For example, students may employ different strategies when preparing for an exam, as opposed to reading for a class assignment (Bråten and Samuelstuen 2007). In a similar vein, students may need to employ different strategies for a mathematics course, as opposed to a humanities course. Furthermore, as described above, most SRL models view SRL as a dynamic, adaptive process. Students' motivation and use of learning strategies may fluctuate over the course of learning (Moos and Azevedo 2008). Although it is possible to account for this by applying short micro-analytic questionnaires at various points during the learning process (Cleary et al. 2015), most common self-report questionnaires are not suited for this purpose. Measurement methods that treat SRL as a static trait, despite the dynamic nature of the underlying models, are considered to be not sensitive enough to these subtle changes in students' SRL. Important information may be lost as a result, making it impossible to answer research questions involving fluctuations in students' SRL strategies within and across learning tasks, and interactions with learner and context characteristics (McCardle and Hadwin 2015).

The second concern regards the issue whether students have the capacity to self-report their use of self-regulatory strategies. Traditional self-report measures of SRL require students to retrieve information about their strategy use from their long-term memory. This can be problematic for four reasons. Firstly, students are likely to have imperfect memory and these memory deficits may cause them to incorrectly report their strategy use. They may overrate the incidence of common events, while underrating the incidence of rare events (Perry and Winne 2006; Tourangeau et al. 2000). Additionally, it is possible that some SRL processes occur subconsciously, leaving students unaware of using them (Perry and Winne 2006). Secondly, the general nature of most self-report questionnaires may leave students uncertain about the context from which to draw the report of their strategy use. This may lead different students to use different contexts when answering the same questionnaire, or individual students to confuse several contexts in which they applied different strategies (Perry and Winne 2006; Schellings 2011). Thirdly, the structured nature of self-report questionnaire items may lead to a situation where students indicate their perceived value of a strategy reflected in a questionnaire item, rather than their actual use of this strategy (Bernacki et al. 2012; Bråten and Samuelstuen 2007). The fourth problem is that students may be inclined to provide socially desirable answers, reporting strategies that they think will please their parents, teachers or the researcher (Bråten and Samuelstuen 2007). In addition to this social desirability, students' ability to report their own use of strategies may also be influenced by how familiar they are with the strategies in the questionnaire. Specifically, learners who have insufficient declarative knowledge about self-regulatory strategies may incorrectly label the strategies they report using (Veenman 2011).

As a result of these issues, researchers have increasingly advocated the use of other, online measures of SRL, in order to adopt a multi-method approach (Veenman 2005; Winne 2010). Examples include systematic observations in which the researcher observes students' outward behavior using a systematic, structured observation instrument (Perry 1998), think-aloud protocols which require students to verbalize their thoughts while working on a task (Ericsson 2006), and trace data in which time-stamped log files are created displaying students' actions in an online learning environment (Winne 2010).

The theoretical background outlined above led us to formulate the following two research questions: 1) How do offline self-report questionnaires compare to online forms of measurement in terms of calibration of students' self-report of strategy use, versus their actual strategy use? 2) Does the degree of calibration vary as a function of the granularity at which SRL is measured? Although SRL is important for students at all educational levels, this review focuses on studies conducted with students in higher education. This decision was made based on two reasons. First, research has shown that the nature and development of SRL is very different for individuals of different age groups (e.g. Schneider 2008). Second, as already hinted at above, the nature of learning in higher education is different than in earlier levels of education, with more demands being placed on students in terms of information seeking and independence. As a result, it would be ill-suited to make the comparison over different age groups at this point. We chose to do a narrative review rather than a meta-analysis for two related reasons. First, very few studies have addressed the measurement of SRL in such a manner that a comparison can be made between students' offline self-reports versus an online form of measurement. Second, in several cases where this comparison was possible, it was not the explicit goal of the research, but rather a byproduct of careful triangulation, thereby giving no statistics or effect sizes for the actual calibration. As a result of these factors, a proper meta-analysis with appropriate effect sizes and sufficient power (Pigott 2012) may not be possible, and a more narrative approach is warranted.

Methods

Search strategy

We conducted a search for English, peer-reviewed articles in the following databases: BioMed Central, ERIC, Medline, PsycInfo, Web of Science and PubMed using the search terms *self-regulated learning calibration*, *(self-regulated learning) AND measur**, and *metacognit* AND measur* AND higher education*. Different combinations of these search terms yielded similar results. Furthermore, the reference lists from the included articles were screened for other relevant publications. We limited our search to include articles between January 2000–May 2016.

The initial search was conducted by the first author, yielding a total of 2059 hits. Based on initial title screening, 580 unique studies were included for abstract screening. After this, fifty-one studies were selected for potential inclusion. These studies were screened by two independent raters. Disagreements regarding inclusion versus exclusion of a study were resolved through discussion. We only included articles which focused on original research with students from higher education, which used and/or compared both offline self-report and an online SRL assessment tool to make a comparison between these different measurements. This resulted in the final inclusion of 14 studies.

Quality assessment

Buckley and colleagues (2009) recommend the following quality criteria on which to base judgments about whether or not to include studies in a review: (1) Does the article provide a clear indication of the *research questions* and hypotheses of the study? (2) Are the *study participants* suitable for the specific study in terms of sample size, selection method,

participant characteristics and homogeneity? (3) Have the researchers used valid and reliable *data collection methods*? (4) Judgment regarding *completeness of data*. How many participants have dropped out of the study? Specifically, the study should have less than 50% attrition, or a response rate of at least 60% in case of survey-based studies. (5) Have the authors applied an appropriate *control for confounding*, accounting for or removing confounding variables if possible? (6) Is the *analysis of results* (statistical or otherwise) appropriate? (7) Do the data provide support for the *conclusions* drawn by the researchers? (8) Is sufficient information provided in the article to enable *reproducibility* of the study? (9) Does the article concern a *prospective* rather than a retrospective study? (10) Did the authors attend to all the *ethical issues* relevant to the study? (11) Was *triangulation* applied, supporting the results with data from multiple sources? In order for a study to be considered of high quality, Buckley et al. (2009) suggest that at least 7 of the 11 quality criteria must be met.

The first and second author independently judged the quality of the studies that were included. A three-point scale was used to judge quality on each of the criteria (+, ±, -). Disagreements were resolved through discussion. On the basis of these quality criteria, all studies were retained in the review. Table 1 summarizes these quality criteria.

Results

When reviewing the literature, we found that in terms of granularity, a general distinction can be made between studies that measure and compare the use of specific self-regulatory strategies, such as highlighting, note creation (fine grained), for at least one of the measures, versus studies measuring a global degree of self-regulatory strategy use (coarse grained), using total scores for self-regulatory activity or scales. Making this distinction led to different conclusions in terms of calibration, as described further below. We will first describe the studies comparing specific strategies, followed by a description of the studies comparing students' global level of self-regulation. Tables 2 and 3 provide an overview of the findings of this review, separated by method of comparison.

Comparison of specific strategies

Ten studies were retrieved that made a comparison between self-reports and online measurements in terms of students' use of specific SRL strategies. These studies can be clustered according to the form of online measurement that was used. We will discuss seven studies using trace data (with four studies focusing on specific learning strategies, and three other studies using goal theory as a starting point), one study using think-aloud protocols, one study using eye movements, and one study online forms of self-report, respectively.

One of the first studies since 2000 to compare offline self-report data with an online measure was conducted by Winne and Jamieson-Noel (2002). These researchers used traces of students' behavior in a software program called PrepMate as an online measure to study the degree of calibration in terms of students' achievement (alignment between students' prediction of achievement and their actual achievement) and self-report of study tactics (alignment between self-reports and traces of study tactics). Students studied a chapter on lightning formation, with achievement being measured using six items addressing all levels of Bloom's taxonomy (Bloom et al. 1956). Questions were worth either five or ten points. After answering a question, students were asked how many of these points they would give themselves, based

Table 1 Quality assessment of the studies included in the review

Reference	Research question	Study participants	Data collection methods	Completeness of data	Control for confounding	Analysis of results	Conclusions	Reproducibility	Prospective	Ethical issues	Triangulation
Adesope et al. (2015)	+	+	+/-	+	+	+	+	+	+	+	+
Bernacki et al. (2012)	+	+	+	+	+	+	+	+	+	+/-	+
Cleary et al. (2015)	+	+	+	+	+	+	+	+	+	+	+
Cooper et al. (2008)	+	+/-	+	+/-	+	+	+	+	+	+	+
De Baeker et al. (2012)	+	+	+	+	+/-	+/-	+	+	+	+/-	+
Hadwin et al. (2004)	+/-	+	+	+	-	+	+	+	+	+/-	+
Hadwin et al. (2007)	+	+	+	+	+	+	+	+	+	+/-	+
Jamieson-Noel and Winne (2003)	+	+	+	+	+	+	+	+	+	+/-	+
Nguyen and Ikeda (2015)	+	+	+/-	+	-	+/-	+	+	+	+/-	+
Sandi-Urena et al. (2011)	+	+	+	+	+	+	+	+	+	+	+
Susac et al. (2014)	+	+	+	+	+	+	+	+	+	+/-	+
Wang (2015)	+	+	+	+	+	+	+	+	+	+/-	+
Winne and Jamieson-Noel (2002)	+	+	+	+	+	+	+	+	+	+/-	+
Zhou and Winne (2012)	+	+	+	+	+	+	+	+	+	+/-	+

Table 2 Schematic overview of studies comparing specific self-regulatory strategies, with + indicating high calibration, – indicating low calibration, and +/- indicating mixed results

Reference	Online	Offline	Type of task	Calibration?	Achievement predicted by?
Adesope et al. (2015)	gStudy	MSLQ, Goal Orientation Questionnaire (GOQ)	Text processing	+/-	n.a.
Bemacki et al. (2012)	nStudy	Previous studies	Text comprehension	+/-	Online/offline
Cleary et al. (2015)	Self-report micro-analytic questions	MSLQ	Exam preparation	-	Online
De Backer et al. (2012)	Think-aloud	Metacognitive Awareness Inventory (MAI)	Critical thinking, problem solving, negotiating, decision making	-	n.a.
Hadwin et al. (2004)	CoNoteS2	Weekly self-report reflections	Text recall and comprehension	-	Online/offline
Hadwin et al. (2007)	gStudy	MSLQ	Text processing for final exam	-	n.a.
Jamieson-Noel and Winne (2003)	PrepMate	Study tactics questionnaire (researcher-developed), estimates of achievement	Text processing (all levels of Bloom's taxonomy)	-	Online/offline
Susac et al. (2014)	Eye-tracking (number of fixations)	Self-report questionnaire (researcher-developed)	Rearranging algebraic equations	-	Online (offline not measured)
Winne and Jamieson-Noel (2002)	PrepMate	Study tactics questionnaire (researcher-developed), estimates of achievement	Text processing (all levels of Bloom's taxonomy)	-	n.a.
Zhou and Winne (2012)	gStudy	Achievement Goal Questionnaire (ACQ)	Text recall and comprehension	-	Online

Table 3 Schematic overview of studies comparing global strategy use, with + indicating high calibration, – indicating low calibration, and +/- indicating mixed results

Reference	Online	Offline	Type of task	Calibration?	Achievement predicted by?
Cooper et al. (2008)	IMMEX	Prospective self-report (Metacognitive Activities Inventory; MCAI)	Chemistry problem-solving	+	Online/offline
Nguyen and Ikeda (2015)	Trace data	MSLQ	University courses	+	n.a.
Sandi-Urena et al. (2011)	IMMEX	Prospective self-report (Metacognitive Activities Inventory; MCAI)	Chemistry problem-solving	–	Offline
Wang (2015)	Think-aloud	Inventory of Metacognitive Self-Regulation (IMSR), confidence judgments	Chemistry problem-solving	+	Online/Offline

on the answer they provided to this question. The self-report questionnaire asked students in how many of the seven paragraphs of the text they had used the respective study strategies (for a full list of strategies, see Winne and Jamieson-Noel 2002). Two items on planning were measured dichotomously and scored as no-planning = 0 or planning = 7. Calibration in study tactics was measured by comparing students' responses on the questionnaire to their behavior in PrepMate, making a comparison between the number of paragraphs in which students reported using the specific study strategies, versus the number of paragraphs in which they were shown to have used these strategies in PrepMate. It was found that despite a consistent general tendency for overconfidence, students were quite well calibrated in terms of their achievement, with a median calibration of $r = .88$ (although quite some variability among different items was found). More importantly however, there was a higher degree of bias and low calibration in students' reporting of their use of study tactics, with a median calibration of $r = .34$. Lowest calibration was found for students' reports of setting objectives and planning a method for the learning task. Furthermore, calibration of study tactics was not related to achievement, while prior knowledge and calibration of achievement were in fact related to achievement. In other words, the degree to which students were able to accurately report their use of study tactics was not related to achievement, but students with higher achievement scores were better able to predict their achievement, when compared to lower achieving students. This indicates that these two forms of calibration tap into different constructs. Prior knowledge was not related to either form of calibration.

In a follow-up analysis, Jamieson-Noel and Winne (2003) again found significant differences between students' self-reports of their study tactics and traces of their actual studying behavior. To investigate the predictive value of traces and self-reports on achievement, separate regression analyses were run for both measurement types. Interestingly, when constructing a measure of traces and self-reported overall SRL intensity by averaging the trace scores and responses to the self-report items respectively, results showed that self-reported SRL intensity (i.e. perceived effort spent with the application of study tactics) significantly predicted achievement (explaining 16% of the variance in achievement), while no contribution was found for traces. After clustering strategies to reflect the planning and learning phases in

Winne and Hadwin's (1998) model of SRL (planning, learning, reviewing and monitoring), traces again did not predict students' achievement. For self-reported strategies, the monitoring phase did emerge as a significant predictor of achievement, explaining 23% of the variance in achievement. It is however important to note that there was no trace for the phase of monitoring, making it impossible for this phase to emerge as a traced predictor. When examining individual tactics, amount of note taking (operationalized as the number of paragraphs in which a student created at least one note) was the only tactic that was a significant predictor when using trace data (23% of variance in achievement explained). In the analysis of self-reports, the significant predictors were reviewing text and review of pictures. In a final analysis, the authors entered both the traces and the self-report items in one blocked regression analysis. In this analysis, the trace for amount of note taking remained a significant predictor of achievement, as well as the self-report items for reviewing text and reviewing pictures, explaining 17% and 26% of the variance in achievement, respectively. Principal component analyses also indicated that traces reveal different forms of SRL than self-reports, with trace data indicating a more active way of studying.

Another study that analyzed students' online traces was conducted by Hadwin et al. (2007). Hadwin et al. (2007) used a similar software program called gStudy to compare eight students' self-reports of self-regulated learning strategies on the MSLQ to their actual use of specific self-regulatory strategies as measured by the traces. Students studied a chapter in a course on introductory educational psychology, which would later be tested on a final exam (no information is given about the content of this exam or students' achievement on this). They clustered students based on their responses to the MSLQ into High, Medium and Low self-regulators. They then tried to identify similarities within clusters in terms of traced study activities. It was found that there were few similarities between students within the same clusters (with even the most highly calibrated students showing good calibration on only 40% of studying activities), indicating that there may be a low calibration between students' self-reports and their actual use of self-regulated learning strategies.

Finally, a study using online traces was conducted by Hadwin et al. (2004), who clustered eight students into the categories of High, Average, Low and Improved performers on the basis of their progression in test performance achievement from pretest to posttest. A software program called CoNoteS2 was used to collect traces of students' studying activities while studying three chapters on sex differences in the context of an instructional psychology course. These trace data were compared with weekly self-report reflections that students wrote regarding their studying tactics. Achievement was measured by students' recall at three levels (unistructural, multistructural and relational), thereby essentially covering text recall and comprehension. They found that High performers were better calibrated than Low performers. However, they also found that studying activities as identified by traces could not independently explain the students' performance developments, indicating the need for additional measures to come to a complete picture.

Zhou and Winne (2012) investigated calibration of a different aspect of SRL, focusing on the comparison of specific achievement goals as measured by self-reports versus trace data. Self-report data were collected with the Achievement Goal Questionnaire (AGQ; Elliot and McGregor 2001). Trace data were collected in gStudy (Winne et al. 2006). In gStudy, participants studied an article about hypnosis, in which they were presented with a predefined set of hyperlinks and tags related to each of the four goal orientations (e.g. "I want to learn more about this" as an indicator of a mastery-approach goal). Goal orientations were inferred by counting the number of hyperlinks students clicked and the number of tags they used.

Achievement was operationalized as text recall and text comprehension. For all goal orientations, there were significant differences between students' self-reports of their goal orientations and the traces collected in gStudy, with effect sizes ranging between $d = 1.39$ and $d = 3.94$. A significant correlation with reading achievement posttest performance was found for traced goal orientations (correlation coefficients ranging between $r_T = .17$ and $r_T = .23$), but not for self-reports.

Also focusing on goal theory, Adesope et al. (2015) investigated whether achievement goals could influence the use of learning strategies, and whether these learning strategies could in turn influence students' online learning behavior. The authors used the Goal Orientation Questionnaire (GOQ; Nesbit et al. 2008) to measure students' goal orientation. Learning strategies were measured using the MSLQ. Students' learning behavior was measured while studying an electronic chapter in gStudy (Winne et al. 2006). Although trace data were used in addition to the self-report questionnaire rather than the two measures being explicitly compared, it is interesting to note that there was a predictive relationship between the questionnaire subscales and learning behavior. Specifically, effort regulation and task value, as measured by the MSLQ, showed a positive predictive relationship with the number of notes and tags that were created in gStudy, as well as with duration of study and the total number of actions completed in gStudy. Furthermore, except for rehearsal, the different learning strategies measured by the MSLQ (elaboration, organization, and metacognitive self-regulation) showed positive correlations with learning behavior, with elaboration showing positive correlation with study duration, the total number of actions, and the total number of notes and tags created, organization showing positive correlations with the total number of actions and the total number of notes and tags created, and metacognitive self-regulation showing a positive correlation with the total number of actions and the number of tags created. Correlation coefficients ranged between $r = .21$ and $r = .42$. This predictive relationship between self-reported learning strategies and students' actual behavior indicates that the MSLQ does in fact tap into an important construct and that students might actually be relatively successful in reporting their use or the importance they assign to these strategies.

Finally, Bernacki et al. (2012) used a trace methodology to examine possible relationships between students' achievement goals, strategy use and comprehension performance. Although they did not make an explicit comparison between traces and self-reports in this study, a comparison was made to earlier studies answering the same research questions using self-report questionnaires. Students used nStudy to study texts on human development and ADHD, with achievement being operationalized as text comprehension. Goal orientation was measured with the Achievement Goals Questionnaire-Revised (Elliot and Murayama 2008). Trace data only replicated a portion of the relationships between goal orientations and learning strategies that were previously reported in self-report studies. Specifically, performance approach goals did not predict any learning strategies, while mastery goals predicted strategies associated with organization and elaboration (specifically note taking and information seeking), and marginally predicted metacognitive monitoring (specifically monitoring of progress), with effect sizes ranging between .13 and 2.75, leaving a general pathway from mastery goals to strategies. Performance avoidance orientation showed a negative relationship with note taking and information seeking behavior, with effect sizes of -1.34 and $-.31$, respectively. The results indicate incongruence between self-reports and trace data for goal orientations, calling into question the validity of self-reports for the measurement of this metacognitive construct.

Situation model comprehension (but not text based comprehension) was predicted by traces of highlighting and progress evaluation, with effect sizes of .05 and .06, respectively.

Furthermore, self-reports of SRL were compared with think-aloud protocols. De Backer et al. (2012) used the prospective Metacognitive Awareness Inventory (MAI; Schraw and Dennison 1994) and a think-aloud protocol to investigate the effect of a reciprocal peer tutoring intervention on students' metacognitive knowledge and strategy use. Students worked on authentic assignments in the context of instructional sciences, requiring critical thinking, problem solving, negotiating and decision making. The questionnaire data and think-aloud protocols showed diverging results. While MAI scores revealed no difference in metacognitive knowledge and regulation between pretest and posttest, think-aloud data showed an increase in the frequency of use of metacognitive skills, with effect sizes ranging between $d = .45$ and $d = 3.12$, as well as an increase in the variation of metacognitive skills.

Furthermore, we found one study that used eye movements as the online measure when making the comparison with offline self-reports. Susac et al. (2014) used eye-tracking data to study students' strategies when rearranging algebraic equations. Eye-tracking data were compared to a self-report questionnaire in which students had to indicate which strategies they had used during the task. Results indicated incongruence between students' self-reports and eye-tracking data. Eye-tracking scan paths revealed several strategies that students did not report in the self-report questionnaire. For example, of the 15 students who indicated that they never checked the provided answers, 51.5% of trials in fact showed a return in eye movements to the answers. In other words, students' metacognitive calibration appeared to be limited, although considerable individual variability was found. Participants who showed higher accuracy in their metacognitive judgments were more successful in efficient equation solving, when compared to students with lower metacognitive accuracy. Furthermore, the eye-tracking data provided a more reliable prediction of equation difficulty, when compared with students' self-reported difficulty rankings. Finally, these eye-tracking measures predicted students' performance in terms of inverse efficiency. Inverse efficiency was operationalized as the ratio between response time and accuracy. Low efficient students showing a higher number returns from answers back to equations than high efficient students, a result which the authors explained by suggesting that high efficient students had better insight into where they should looking, thereby requiring fewer returns. However, the authors did not compare this result to the questionnaire data.

Finally, some studies have compared the use of offline self-report questionnaires to online forms of self-report. Cleary et al. (2015) compared students' responses to the MSLQ to their responses to self-report micro-analytic questions delivered to the students by the examiner, assessing exam preparation. The relationship between students' MSLQ scores and their responses to the micro-analytic strategy questions was not significant. Furthermore, the micro-analytic strategy questions were a better predictor of students' academic performance than the MSLQ. Specifically, there were no significant correlations between exam scores and MSLQ scales, while the weighted micro-analytic strategy measure significantly predicted students' grade on the final exam, with a correlation coefficient of $r = .29$.

Overall, studies that focus on the use of specific strategies when comparing self-report questionnaires with behavioral measures indicate low calibration between the two forms of measurement (Adesope et al. 2015; Bernacki et al. 2012; Cleary et al. 2015; De Backer et al. 2012; Hadwin et al. 2004, 2007; Jamieson-Noel and Winne 2003; Susac et al. 2014; Winne and Jamieson-Noel 2002; Zhou and Winne 2012). Traces tend to have a higher predictive value in terms of achievement than self-reports.

Comparison of global use of self-regulatory strategies

As opposed to the 10 studies comparing different types of measurement for specific self-regulatory strategies, four other studies have focused on a global measure of self-regulation, using total or subscale scores that aggregate different self-regulatory strategies. Three studies focused on problem-solving, while one study used an electronic portfolio system.

Cooper et al. (2008) developed a multi-method instrument to measure students' metacognition in chemistry problem-solving across time. In order to do so, they compared students' answers on the prospective self-report Metacognitive Activities Inventory (MCA-I; Cooper and Sandi-Urena 2009) to their study strategies in an online problem-solving environment called IMMEX. In IMMEX, students work on ill-defined chemistry problems while their problem-solving activities are recorded. For example, the number of relevant information pieces considered before trying to solve a problem is used as an indicator of planning. The researchers found convergence between their self-report instrument and students' behavior in the online environment, in the sense that students who performed more metacognitive strategies in the online environment also had higher scores on the questionnaire, as compared to students who executed fewer metacognitive strategies. Furthermore, there was a significant correlation between students' problem-solving performance and their strategy use in the online environment, as well as with their scores in the self-report questionnaire.

In a later study on problem-solving, Sandi-Urena et al. (2011) used the MCA-I and the IMMEX environment to assess the effects of a cooperative intervention on students' metacognitive awareness and strategy use. In this study, the intervention led to a *decrease* in self-reported metacognitive strategy use (interpreted by the authors as an increase in metacognitive awareness) as measured by the MCA-I (with an effect size of $d = .10$ for the difference between the two groups at posttest), but no changes were observed in actual use of metacognitive strategies in the IMMEX environment. Regardless of the direction of the results and the interpretation of this (a decrease in metacognitive strategies versus an increase in metacognitive awareness), the inconsistency between the self-report questionnaire and the use of metacognitive strategies in the IMMEX environment points to an incongruence between students' self-report and the trace data. As an explanation for this incongruence, the authors propose that the MCA-I might put a greater emphasis on reflection, rather than metacognitive skill application. However, we propose it could also be due to a greater sensitivity of the MCA-I to changes from pretest to posttest, or a lower validity of this instrument with students reporting socially desirable answers as a result of having been exposed to the intervention. Interestingly though, the intervention did lead to an increase in students' problem-solving ability, suggesting that the increase in MCA-I scores might have tapped into an actual change in students' strategies.

Finally, Wang (2015) used a multimethod approach to investigate the general and task-specific aspects of metacognition in different topics in chemistry problem solving (molecular polarity and thermodynamics). Self-reported metacognitive skill was measured with the Inventory of Metacognitive Self-Regulation (IMSR; Howard et al. 2000). Concurrent metacognitive skill was measured using a think-aloud protocol. Furthermore, confidence judgments and calibration accuracy values were obtained. Results indicated a significant association between self-report questionnaire scores and concurrent metacognitive skill use as measured by the think-aloud protocol (with a correlation coefficient of $r = .36$ for the thermodynamics task, and $r = .49$ for the molecular polarity task). For the task on molecular polarity, both the self-report questionnaire and the think-aloud protocols showed a significant

correlation with performance ($r = .39$ and $r = .55$, respectively). For the thermodynamics task, only the think-aloud protocols showed a significant correlation with performance (with a correlation coefficient of $r = .40$). The author concludes that the self-report questionnaire assesses a context-independent, general and common aspect of metacognition, while think-aloud methodology assesses context-specific metacognition.

Nguyen and Ikeda (2015) developed and evaluated an electronic portfolio system to support SRL in students in the context of two university courses with ICT topics. They used the MSLQ to measure self-reported SRL strategies and examined traces in the ePortfolio environment to assess students' actual use of strategies. Results indicated differences from pretest to posttest and between experimental groups for MSLQ scores, congruent with overall increases in SRL strategies observed in the trace data, which could be interpreted as calibration of self-reported study strategies.

Taken together, these studies (Cooper et al. 2008; Nguyen and Ikeda 2015; Sandi-Urena et al. 2011; Wang 2015) indicate that when studies examine the global level of self-regulation, students are relatively well able to report on their use of self-regulatory strategies. This is in contrast with the results from the studies comparing specific self-regulatory strategies, where low calibration is found between the two types of measurement. There appears to be individual value of self-reports of global self-regulation when predicting academic achievement. Self-reports of strategy use can predict achievement, over and above the predictive value of the trace data that were used in the studies. These differential results indicate that different types of measurement (self-report versus online measures) are appropriate for different types of research questions or interventions, a point further elaborated upon in the Discussion.

Discussion

In this review, we compared offline self-report questionnaires with online behavioral instruments to assess self-regulated learning. Granularity was found to be an important construct when it comes to the comparison between offline self-reports and online measurements, influencing the level of convergence between students' self-reports and behavioral indicators of SRL. Studies that indicate high calibration are mainly those with a focus on students' global use of self-regulatory strategies (coarse grained). Studies that focus on calibration of concrete self-regulatory strategies (fine grained) generally indicate a low degree of calibration. Apparently, students are able to report on overall degree, increase or decrease of their use of self-regulatory strategies in general, indicating calibration when SRL is measured at this coarse grain size. However, they have difficulty pinpointing the exact strategies they use when SRL is measured at a fine grain size. Depending on the researcher's specific research question or problem statement this may or may not be a problem. For example, when creating an intervention to increase students' global metacognitive awareness, it might be sufficient to measure this with a self-report questionnaire. Furthermore, in order for interventions to be effective, it is important to also take into account the students' perceptions about their own self-regulatory abilities (Perry and Rahim 2011). Self-report can play an important role in this regard. However, when focusing on the development of specific self-regulatory strategies, for example in the context of an intervention to develop deep learning strategies in students, the use of more online measures of SRL (e.g. trace data) might be warranted. If a link should be made to learning outcomes, trace data have been found to be a more powerful predictor.

However, as with any form of measurement, it is important that the online measurements represent a valid way of assessing SRL. When using online measures of SRL, the operationalization of strategies is an important consideration. Both computerized log data (Winne 2010) and measures such as eye-tracking (Kok and Jarodzka 2016) are meaningless without the use of a sound underlying theoretical model. It is important to realize that behavioral measures may obscure mental operations, which could in fact be captured by self-report. For example, Winne and Jamieson-Noel (2002) defined students' planning of a method as scrolling through the text before performing any of the other traced learning strategies. They found that method planning was one of the strategies in which students were especially poorly calibrated. While they used a very plausible operationalization, it is also very well possible this planning occurred entirely in students' heads. In fact, a notable exception to the general finding that traces have a higher predictive value for achievement than self-reports is the study by Jamieson-Noel and Winne (2003), in which traces of SRL intensity did not predict achievement, while self-report of SRL intensity did. Clusters of traces reflecting the different phases in Winne and Hadwin's (1998) SRL model also did not predict achievement, while self-reports of monitoring did. This might be due to the fact that in this analysis, traces were again clustered to reflect global scales. In fact, when zooming in on individual strategies, some traces did predict achievement. Also, the study included no traces for monitoring, making it impossible to find a traced effect of this scale on students' achievement. In order for research on self-regulated learning and calibration to advance further, there is a need for agreement on an overarching framework of SRL strategies and how to measure them in electronic learning environments or other behavioral measures. In order for this to be possible, a firm theoretical grounding is important. Measures of SRL should be closely aligned to their underlying models, which in the reported studies is often not the case.

The distinction between measurement of global SRL versus specific strategies could also explain why the behavioral measures of SRL (traces, think-aloud protocols, micro-analytic questions and eye-tracking data) tend to be better predictors of academic achievement than students' self-reports. It is conceivable that achievement can be predicted by some strategies (the "good" ones) but not by others, precluding the predictive value of measures of "global" self-regulation, as these tend to aggregate students' responses over multiple occasions and combine multiple different strategies into a few subscales. For example, it has been found that deep strategies such as elaboration and organization are more effective than more shallow strategies such as rehearsal (Pintrich et al. 1993), and even within these categories, some strategies are likely to be more effective than others. For example, when studying materials focusing on connections between constructs it might be helpful to use organization strategies, while strategies related to elaboration might be more appropriate for grasping global theories and systems. Furthermore, different strategies may be more effective and/or better calibrated at different phases of learning (Greene and Azevedo 2007). Methods that aggregate such strategies are likely to obscure potential effects on achievement. Future research should focus on clearly delineating the predictive effect of individual strategies on specific learning tasks in higher education, in order to further inform interventions to enhance SRL in students.

In recent years, researchers have also emphasized the social aspects involved in SRL (e.g., Hadwin et al. 2011; Hadwin and Oshige 2011; Järvelä and Hadwin 2013). Rather than treating the social context as one of the components in the SRL process, these researchers place shared knowledge construction at the center of learning (Hadwin et al. 2011). This perspective has implications for the way in which SRL should be measured. Specifically, measurements should be used that are able to capture this reciprocity (Hadwin et al. 2010), without ignoring

the temporal and sequential aspects of the interactions (Molenaar and Järvelä 2014). The trace data measurements outlined in this review can play an important role in such research (Hadwin et al. 2010). When properly designed, they can offer an efficient, highly detailed alternative to traditional classroom observations. Related to this point, future research could also examine to what extent the accuracy of students' self-reports of strategy changes in an isolated individual versus a social context.

Some weaknesses should be noted. An interesting finding that emerged from the review is that the level of convergence between students' self-reports and behavioral indications of SRL depends on the granularity of measurement. When comparing specific self-regulatory strategies, students seem unable to self-report on their strategy use. Conversely, when comparing global self-regulation, a higher level of convergence is found. However, only few studies have set out to compare the use of self-report questionnaires to more online forms of measurement. The number of studies that have focused on global self-regulation, as opposed to specific strategy use, has been particularly small. Sample sizes are sometimes small and many of the studies have been conducted by the same groups of researchers. These considerations led us to conduct a narrative, rather than a systematic review. Consequently, the field could be further advanced by more research by different groups of researchers in different populations of students, in order to replicate the results found in this review. In the future, these studies could be synthesized into a more systematic review or meta-analysis of the literature, providing clearer insight into the individual value of both self-reports and behavioral indicators of SRL. Finally, we have focused this review on SRL strategies in students in higher education. It can be expected that there will be differences in SRL and calibration between different age groups. It would be interesting to make this comparison for other age groups as well.

Furthermore, the studies included in this review suffer from another weakness inherent in the use of self-report. Specifically, without an external criterion of self-regulation, it is difficult to establish whether self-regulation has occurred in the first place. We have attempted to mitigate this problem in this review by only including studies that made a comparison with an online form of measurement, but since online measurements also require considerable operationalization and interpretation, we can never be entirely sure about the nature of the constructs being compared. This issue highlights the importance of properly triangulating measures in research on self-regulated learning.

Finally, the studies described in this review employed several different task types (problem solving, text comprehension, etc.). To our knowledge, research has not focused on how the overlap between students' self-report versus online measures of their self-regulation might differ according to task type, which is surprising given the fact the literature does indicate differences in self-regulation according to the context (McCardle and Hadwin 2015; Winne and Hadwin 1998; Bråten and Samuelstuen 2007). It fell beyond the scope of this review to go into a detailed comparison according to task type, but this could be a fruitful area for further research.

The main conclusion that can be drawn from this review is that self-report questionnaires have their own value in educational research and remediation, in the sense that they might give a relatively accurate insight into students' global level of metacognition, serving as a starting point for more precise interventions. Furthermore, when students' perceptions of their self-regulation are the focus, self-reports can be instrumental in providing this insight (Perry and Rahim 2011). What matters is that researchers and educationalists think carefully about the research questions or problems they wish to address, being aware of the affordances and limitations of different measurement methods,

and align their measurements to the issues at hand. Although these conclusions and implications are not highly specific, this observation provides us with important information about the state-of-the-art of research in this field. As Winne (2017) states: “Because expressions of metacognition in SRL are complex, research upon which to base practice may appear piecemeal, failing to paint a whole picture” (p. 45). We hope that this review can be a first step in the direction of a more complete picture.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Adesope, O. O., Zhou, M. M., & Nesbit, J. C. (2015). Achievement goal orientations and self-reported study strategies as predictors of online studying activities. *Journal of Educational Computing Research*, 53(3), 436–458. <https://doi.org/10.1177/0735633115603989>.
- Arbesman, S. (2013). *The half-life of facts. Why everything we know has an expiration date*. New York: Penguin Group.
- Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: a discussion. *Metacognition and Learning*, 4(1), 87–95.
- Bernacki, M. L., Byrnes, J. P., & Cromley, J. G. (2012). The effects of achievement goals and self-regulated learning behaviors on reading comprehension in technology-enhanced learning environments. *Contemporary Educational Psychology*, 37(2), 148–161. <https://doi.org/10.1016/j.cedpsych.2011.12.001>.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Kratochwill, D. R. (1956). *Taxonomy of educational objectives. Handbook 1: Cognitive domain*. New York: McKay.
- Bråten, I., & Samuelstuen, M. S. (2007). Measuring strategic processing: comparing task-specific self-reports to traces. *Metacognition and Learning*, 2(1), 1–20.
- Cleary, T. J., Callan, G. L., Malatesta, J., & Adams, T. (2015). Examining the level of convergence among self-regulated learning microanalytic processes, achievement, and a self-report questionnaire. *Journal of Psychoeducational Assessment*, 33(5), 439–450. <https://doi.org/10.1177/0734282915594739>.
- Cooper, M. M., & Sandi-Urena, S. (2009). Design and validation of an instrument to assess metacognitive skillfulness in chemistry problem solving. *Journal of Chemical Education*, 86(2), 240.
- Cooper, M. M., Sandi-Urena, S., & Stevens, R. (2008). Reliable multi method assessment of metacognition use in chemistry problem solving. *Chemistry Education Research and Practice*, 9(1), 18–24. <https://doi.org/10.1039/b801287n>.
- De Backer, L., Van Keer, H., & Valcke, M. (2012). Exploring the potential impact of reciprocal peer tutoring on higher education students’ metacognitive knowledge and regulation. *Instructional Science*, 40(3), 559–588. <https://doi.org/10.1007/s11251-011-9190-5>.
- Elliot, A. J., & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80(3), 501–519.
- Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: critique, illustration, and application. *Journal of Educational Psychology*, 100(3), 613.

- Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. Hoffman. (Eds.), *Handbook of expertise and expert performance*. New York: Cambridge University Press. (pp. 223–241).
- Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: new perspectives and directions. *Review of Educational Research*, 77(3), 334–372.
- Hadwin, A. F., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: exploring perspectives of social in self-regulated learning theory. *Teachers College Record*, 113(2), 240–264.
- Hadwin, A. F., Boutara, L., Knoetzke, T., & Thompson, S. (2004). Cross-case study of self-regulated learning as a series of events. *Educational Research and Evaluation*, 10(4–6), 365–417. <https://doi.org/10.1080/13803610512331383499>.
- 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(2–3), 107–124.
- Hadwin, A. F., Oshige, M., Gress, C. L., & Winne, P. H. (2010). Innovative ways for using gStudy to orchestrate and research social aspects of self-regulated learning. *Computers in Human Behavior*, 26(5), 794–805.
- Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 65–84). New York: Routledge.
- Howard, B. C., McGee, S., Shia, R., & Hong, N. S. (2000). *Metacognitive self-regulation and problem-solving: Expanding the theory base through factor analysis*. Paper presented at the American Educational Research Association, New Orleans, LA.
- Jamieson-Noel, D., & Winne, P. H. (2003). Comparing self-reports to traces of studying behavior as representations of students' studying and achievement. *Zeitschrift Fur Pädagogische Psychologie*, 17(3–4), 159–171. <https://doi.org/10.1024/1010-0625.17.3.159>.
- Järvelä, S., & Hadwin, A. F. (2013). New frontiers: regulating learning in CSCL. *Educational Psychologist*, 48(1), 25–39.
- Kok, E. M., & Jarodzka, H. (2016). Before your very eyes: the value and limitations of eye tracking in medical education. *Medical Education*, 51(1), 114–122.
- McCardle, L., & Hadwin, A. F. (2015). Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. [article]. *Metacognition and Learning*, 10(1), 43–75. <https://doi.org/10.1007/s11409-014-9132-0>.
- Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, 9(2), 75–85.
- Moos, D. C., & Azevedo, R. (2008). Exploring the fluctuation of motivation and use of self-regulatory processes during learning with hypermedia. *Instructional Science*, 36(3), 203–231.
- Nesbit, J. C., Zhou, M., Mahasneh, R., & Yeung, P. (2008). *The goal orientation questionnaire (GOQ)*. Burnaby: Simon Fraser University.
- Nguyen, L. T., & Ikeda, M. (2015). The effects of ePortfolio-based learning model on student self-regulated learning. *Active Learning in Higher Education*, 16(3), 197–209. <https://doi.org/10.1177/1469787415589532>.
- Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology*, 90(4), 715.
- Perry, N. E., & Rahim, A. (2011). Studying self-regulated learning in classrooms. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 122–136). New York: Routledge.
- 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(3), 211–228.
- Pigott, T. (2012). *Advances in meta-analysis*. New York: Springer Science & Business Media.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385–407.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Ann Arbor: National Center for Research to Improve Postsecondary Teaching and Learning, University of Michigan.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, 53(3), 801–813.
- Sandi-Urena, S., Cooper, M. M., & Stevens, R. H. (2011). Enhancement of metacognition use and awareness by means of a collaborative intervention. *International Journal of Science Education*, 33(3), 323–340. <https://doi.org/10.1080/09500690903452922>.
- Schellings, G. L. M. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning*, 6(2), 91–109.

- Schellings, G. L. M., & Van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: theoretical and empirical considerations. *Metacognition and Learning*, 6(2), 83–90.
- Schellings, G. L. M., van Hout-Wolters, B. H., Veenman, M. V., & Meijer, J. (2013). Assessing metacognitive activities: the in-depth comparison of a task-specific questionnaire with think-aloud protocols. *European Journal of Psychology of Education*, 28(3), 963–990.
- Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: major trends and implications for education. *Mind, Brain, and Education*, 2(3), 114–121.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475.
- Susac, A., Bubic, A., Kaponja, J., Planinic, M., & Palmovic, M. (2014). Eye movements reveal students strategies in simple equation solving. *International Journal of Science and Mathematics Education*, 12, 555–577.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- 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. (2011). Alternative assessment of strategy use with self-report instruments: a discussion. *Metacognition and Learning*, 6(2), 205–211.
- 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), 3–14. <https://doi.org/10.1007/s11409-006-6893-0>.
- Wang, C.-Y. (2015). Exploring general versus task-specific assessments of metacognition in university chemistry students: a multitrait-multimethod analysis. *Research in Science Education*, 45, 555–579.
- Weinstein, C. E., & Palmer, D. R. (2002). *Learning and study strategies inventory* (2nd ed.). Clearwater: H & H Publishing.
- Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, 45(4), 267–276. <https://doi.org/10.1080/00461520.2010.517150>.
- Winne, P. H. (2017). Cognition and metacognition within self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance*. New York: Routledge, (2nd ed., pp. 36–48)
- 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. 27–30). Mahwah: Lawrence Erlbaum.
- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27, 551–572.
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). Orlando: Academic Press.
- 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 measured and methods* (Vol. 12, pp. 121–155). Greenwich, Connecticut: JAI.
- Winne, P. H., Nesbit, J. C., Kumar, V., Hadwin, A. F., Lajoie, S. P., Azevedo, R., et al. (2006). Supporting self regulated-learning with gStudy software: the learning kit project. *Technology, Instruction, Cognition and Learning*, 3, 105–113.
- Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, 22, 413–419.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social-cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego: Academic Press.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166–183. <https://doi.org/10.3102/0002831207312909>.
- Zimmerman, B. J., & Schunk, D. H. (1989). *Self-regulated learning and academic achievement: Theory, research, and practice*. New York: Springer.