Instructional Science (2021) 49:137–163
https://doi.org/10.1007/s11251-020-09533-2

ORIGINAL RESEARCH

Multimethod assessment of self-regulated learning in college students: different methods for different components?

Laura Dörrenbächer-Ulrich[1] · Marie Weißenfels[1] · Lea Russer[1] · Franziska Perels[1]

Received: 22 June 2019 / Accepted: 21 December 2020 / Published online: 11 January 2021
© The Author(s) 2021

Abstract
Although self-regulated learning (SRL) is seen as highly relevant for successful college learning, college students oftentimes show a lack in SRL abilities**.** Therefore, it seems necessary to foster SRL in this group of leaners. In order to evaluate such training and to foster SRL in an optimal way, a valid assessment of this competence and its development is necessary. As different methods for the assessment of SRL show benefits and points of criticism, the present study used a multimethod approach to investigate convergence between and across different measures as well as their predictive validity for achievement. SRL was conceptualized of cognitive, metacognitive, and motivational components. Seventy college students were assessed with two broad SRL-measures (questionnaire, strategy knowledge test) and two task-specific SRL measures (microanalyses, trace data) within a standardized laboratory setting. Moreover, GPA of college entrance diploma was gathered as an indicator of general achievement level. Results indicate moderate to high relations between the different components of SRL (cognition, metacognition, and motivation) within one assessment level and no relations between the different assessment methods within one component. With regard to achievement, we found that every component is predictive for achievement but only if measured with different assessment methods. The results are discussed with regard to their implications for future research and the use of different assessment methods for SRL.

Keywords Self-regulated learning · Microanalysis · Strategy knowledge · Trace data · College students

Introduction

College students are confronted with autonomous learning settings that differ clearly from structured learning settings in high school (Cohen 2012). Learning problems can result from this newly acquired autonomy and possible feelings of isolation and struggling (Wei

✉ Laura Dörrenbächer-Ulrich
laura.doerrenbaecher@uni-saarland.de

[1] Saarland University, Campus A4.2, 66123 Saarbrücken, Germany

et al. 2005). As college students are expected to acquire relevant knowledge for their success in a self-guided and self-directed way (Gläser-Zikuda et al. 2010), self-regulated learning (SRL) is seen as highly relevant for successful college learning (Bembenutty 2011). College students oftentimes show a lack in SRL abilities (Peverly et al. 2003), although the usage of SRL strategies is positively related to academic achievement (e.g., Kitsantas 2002). Therefore, college students can especially benefit from SRL strategy trainings (e.g., Bellhäuser et al. 2016; Dörrenbächer and Perels 2016). To evaluate and foster SRL in an optimal way, it is important to know how to assess this competence, as well as its development, adequately. Recent research has suggested different ways of assessing SRL (questionnaires, McCardle and Hadwin 2015; knowledge tests, Händel et al. 2013; microanalysis, Callan and Cleary 2017; trace data, Azevedo et al. 2010), but all methods show benefits as well as points of criticism: For example, some measures are easy to use with large groups of students (e.g. questionnaires), but they oftentimes have to be answered retrospectively, so they are prone to retention or generalization problems. In contrast, more task-specific measures, such as microanalysis, allow for a fine-grained assessment of learning-relevant behavior during the learning opportunities in which they occur, but they present challenges with regard to an objective evaluation of answers. To analyze the convergence between different measures and to compare their correlation with achievement, a multimethod approach seems indispensable (Winne and Perry 2000). As only few studies have used such an approach, the present study investigates the convergence between two broad SRL measures (questionnaire, strategy knowledge test) and two task-specific SRL measures (microanalyses, trace data) analyzing their correlation with academic achievement (GPA) in college students.

Definition and models of SRL

In the context of college students' learning, SRL is often defined as "processes whereby learners personally activate and sustain cognitions, affects, and behaviors that are systematically oriented towards the attainment of personal goals" (Zimmerman 2011, p. 1). There are generally two types of models to describe SRL (Wirth and Leutner 2008): Component models on the one hand specify different levels of regulation and are relatively static. Therefore, they consider SRL as a rather stable competence that is characterized by distinct components. Process models on the other hand describe SRL as a dynamic state that changes depending on environmental and situational conditions (Schmitz and Wiese 2006). This adaption can be ensured through feedback loops that inform the individual about ineffective strategies or unrealistic goals and therefore help optimize learning behavior. Although there are several definitions and models of SRL, most authors agree that it comprises cognitive, metacognitive, and motivational components (Boekaerts 1999). The three-layer model of SRL (Boekaerts 1999) is a component model and focuses on the interaction of these cognitive, metacognitive, and motivational regulatory processes referring to three hierarchical layers. The inner layer comprises the regulation of information processing and concerns the knowledge and effective use of cognitive learning strategies. The second layer refers to the regulation of the whole learning process and focuses on the usage and control of learning strategies. It therefore represents metacognitive knowledge and the use of metacognitive strategies such as planning, monitoring, and reflecting on learning processes in an adaptive way. The outer layer reflects the regulation of the self and therefore encompasses motivational components (e.g. intrinsic motivation, goal setting) as well as motivational beliefs (causal attribution, self-efficacy), and volitional strategies that

support the initiation and maintenance of learning processes. Although the layers interact with each other, the model is relatively stable as it describes different regulatory focuses that require using layer-specific strategies to optimize learning processes.

In contrast to component models, process models are based on the social-cognitive framework (Bandura 1986), which deals with the triadic interaction of person, behavior, and environment and defines self-regulation as "self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals" (Zimmerman 2000, p. 14). Within Zimmerman's model (2000), self-regulation is domain-unspecific as it can be applied to almost all domains such as health behavior and sports. If self-regulation is applied to the context of learning, the construct is named self-regulated *learning* (SRL) and is divided into three cyclical phases of forethought, performance or volitional control, and self-reflection. The basis to initiate the forethought phase of an SRL process is self-motivation. One essential self-motivational belief is self-efficacy (Bandura 1997), which is defined as "personal beliefs about having the means to learn or perform effectively" (Zimmerman 2000, p. 17). Other important components of the forethought phase are the selection of specific goals and the strategic planning of the learning process, as these provide a frame for the following phases of performance or volitional control and self-reflection.

During the phase of performance and volitional control, self-control is necessary for learners to focus on performance and can be supported by cognitive learning strategies for deep processing (e.g., organization and elaboration of the subject matter). Moreover, self-monitoring and volitional control are of special relevance during the performance phase, as they help to observe one`s own progression and emerging difficulties while learning and therefore offer the opportunity to modify or adapt strategies. The following self-reflection phase serves to self-evaluate the learning process relating to different criteria, for example, goal attainment. The learner should try to derive consequences and to implement adaptive self-reactions to improve further learning processes.

Relevance of SRL for college students

Self-regulated learning oftentimes is seen as a crucial part of lifelong learning (Wirth and Leutner 2008). A positive relation between SRL and academic achievement has been found in different learning environments covering the entire educational process (primary school, e.g., Throndsen 2011; secondary school, e.g., Perels et al. 2009; college, e.g., Kitsantas et al. 2008). Although SRL is relevant for all educational settings, it is especially relevant for college students since the transition from high school to college embraces challenges and changes in the learning context as well as in private life (Park et al. 2012). College students experience a high amount of pressure, as they have to cope with extensive learning material within a short period of time (Cohen 2012). Although SRL in general has a high relevance for college student learning, time management strategies (Kitsantas et al. 2008), elaboration strategies (Tynjälä et al. 2005), and self-efficacy (DiBenedetto and Bembenutty 2013; Richardson et al. 2012) have empirically shown to be especially important for academic success. Moreover, a meta-analysis of Sitzmann and Ely (2011) underlines the importance of metacognitive strategies, motivation, attribution, goal setting, and self-efficacy for successful learning processes. With regard to metacognitive strategies, it is especially important for students to estimate their own abilities and knowledge adequately as an overestimation is associated with poorer academic achievement, possibly mediated through a weaker commitment to work (Dunlosky and Rawson 2012).

Indeed, past research shows that SRL is positively correlated with academic achievement in college (Kitsantas 2002; DiBenedetto and Bembenutty 2013), but college students' learning behavior oftentimes does not reflect the ideal usage of SRL strategies (Peverly et al. 2003). For example, some students show difficulties in evaluating the conditions and usage of learning strategies and therefore prefer surface strategies (Schmidt et al., 2011). Consequently, it seems necessary to foster SRL in college students. Several studies have shown that SRL trainings can be effective (Dörrenbächer and Perels 2016; Zimmerman et al. 2011) and show long-lasting effects (Bail et al. 2008). To develop optimal training interventions and evaluate the effectiveness of the training, it is essential to design instruments for the assessment of SRL that are reliable and valid. Therefore, the following section provides an overview of existing instruments concerning their benefits and points of criticism.

Assessment of SRL

Assessment of SRL, thus of cognitive, metacognitive as well as motivational processes, is a field of great interest that has led to many different approaches during the last decades (e.g., Azevedo et al. 2011; Händel et al. 2013; Panadero et al. 2016; Veenman and van Cleef 2018). On a conceptual level, one can distinguish between approaches that aim at considering general tendencies in learning behavior on a global level (*aptitude measures*) and approaches that consider the actual learning situation (*event measures*) on a very specific level (Cleary and Callan 2018). Aptitude measures oftentimes are task-unspecific, whereas event measures mostly are connected to a specific academic task and therefore are task-specific. In the following, we will describe the most common measures and explain the benefits and points of criticism of the measures used in this study (see Table 1).

The most common *aptitude measure* is self-report questionnaires (Roth et al. 2016), such as the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al. 1993). Such questionnaires assess the average use of cognitive, metacognitive, and motivational learning strategies in typical learning situations. As Rovers et al. (2019) clarify, self-report measures have benefits and weaknesses: They are very economic as they can be administered to a large sample of learners in an easy way. Additionally, many different strategies and attitudes can be assessed in a short amount of time (Wolters and Won 2018). Moreover, questionnaires are objective approaches for the assessment of SRL because they are standardized in implementation and interpretation. In contrast to these benefits, there are two main points of criticism of self-report measures (Rovers et al. 2019): First, as SRL is defined as a context-dependent process, measures that are based on a trait definition of SRL (as questionnaires do) are not sensitive enough to capture changes in SRL behavior. Second, it is questionable if students are able to precisely report on their use of SRL strategies. They have to rate their strategy usage retrospectively and aggregated over time and situations, which is prone to retention or generalization problems (Winne and Perry 2000). Moreover, self-report measures are always suspect to social desirability bias. Furthermore, one cannot be aware of all cognitive, metacognitive, and motivational processes that are proceeding during a learning situation, as some of these processes occur subconsciously, which makes an accurate report nearly impossible. In addition, questionnaires are not clear on the context that should be used for evaluating strategy use, so that every student uses another context to answer the questionnaire. Moreover, several authors suspect validity problems of questionnaires, as they mix up the assessment of strategy knowledge and strategy usage and do not reflect on actual behavior (Artelt 2000). Therefore, changes in

Table 1 SRL measures used in this study and their benefits as well as points of criticism

Measurement level	SRL measure	Examples	Benefits	Points of criticism
Global or aptitude level	Questionnaires	MSLQ; Pintrich et al. (1993)	Economic Objective and standardized	Not sensitive to changes in SRL behavior Prone to retention or generalization problems Suspect to social desirable answers No report of subconscious processes possible Unclear on context that should be drawn on Mix up assessment of strategy knowledge and strategy usage
	Strategy knowledge tests	WSLT 7–12; Schlagmüller & Schneider (2007), Maag Merki, Ramseier & Karlen (2013)	Economic Objective and standardized Clear context that should be drawn on Not mixing up strategy knowledge and strategy usage	Not sensitive to changes in SRL behavior Prone to retention or generalization problems Suspect to social desirable answers Unclear if strategy knowledge is a predictor of strategy usage
Situational or event level	Microanalytic assessment	Cleary and Callan (2018)	Fine-grained assessment of learning-relevant behavior during actual learning situations Non-suggestive open questions	Urge for a well-defined learning task Open questions need a clear categorization for assessment Labor-intensive assessment of open questions
	Trace data	Winne (2010)	Precise and objective Help capture the dynamic character of SRL processes	Only give information on observable behavioral traces, therefore not appropriate for every SRL component (e.g. motivational thoughts)

questionnaire values can only be interpreted as shifts in students' perception of their SRL behavior and the extent to which they engage in SRL processes. Despite these weaknesses, McCardle and Hadwin (2015) underline that questionnaires "provide important information for examining and interpreting SRL even when the reports are inaccurate or skewed" (p. 46), as students use their own perception for goal setting, self-monitoring and adapting their behavior. Moreover, Rovers et al. (2019) found that students are able to report on their global level of SRL using questionnaires and that these self-reports can predict achievement. Nevertheless, in studies that try to predict learning outcomes by self-reported SRL strategies, only 4% of variance could be explained (Veenman and Spaans 2005). This prediction can be improved by framing specific situations to ask about strategy usage instead of general strategy usage over many different situations (Leopold and Leutner 2002).

This is where strategy knowledge tests come into play. In these tests, students have to rate the usefulness of different strategy suggestions for specific learning situations and analyses measure how close the rating is to the rating of SRL experts, as well as theoretical assumptions (Händel et al. 2013). This approach is based on the assumption that knowledge on SRL strategies is a prerequisite for successful application of these strategies (Grassinger 2011; Wolters 2003). Most of the developed strategy knowledge tests refer to meta-strategic knowledge according to Kuhn (2000) and are subject-specific, such as reading strategy knowledge tests (*Würzburger Lesestrategiewissenstest*, WSLT 7–12; Schlagmüller and Schneider 2007) that focus on reading strategies. Studies employing this instrument reveal strong correlations with reading competence (Artelt et al. 2010). Comparable results can be found for mathematics (Lingel et al. 2014; Ramm et al. 2006) as well as English (Neuenhaus et al. 2017). Advantages of this approach are comparable to questionnaires and particularly contain benefits related to economy and objectivity. Concerning the SRL framework, tests that measure (metacognitive) strategy knowledge are relatively scarce and primarily exist in German-speaking countries (e.g., Händel et al. 2013, for lower secondary school students; Maag Merki et al. 2013, for upper secondary school students and college students). These tests present different school-specific learning scenarios, each followed by a list of learning strategies that differ "in their degree of effectiveness for the given situation" (Händel et al. 2013, S. 174), whereas the effectiveness was assessed through experts in the field of learning strategies. To compute learning strategy knowledge scores, the students' rating of pairs of strategies (in which one strategy is rated as useful and one as not useful by the experts) is coded. Although these tests assess strategy knowledge, the scenarios presented are not applicable for college students and the underlying theoretical framework is relatively narrow (e.g., measuring only the metacognitive component of SRL).

A more task-specific possibility to assess SRL are microanalytic assessments. They allow for a fine-grained assessment of learning-relevant behavior, cognition or affect during the learning opportunities in which they occur (Cleary 2011) and therefore represent event measures. Microanalytic self-reports, contrary to questionnaires, do not show a retrospective or prospective bias as they are answered during the learning situations itself (Cleary et al. 2012). According to the general model of self-regulation (Zimmerman 2000), it is possible to gather information about the three SRL phases of forethought, performance and volitional control, and self-reflection (Cleary and Callan 2018). Questions about the planning process are asked before learning, questions about the performance phase can be asked during learning, and questions about the self-reflection phase are asked after having finished the learning task. The learning task itself has to be well defined with a clear beginning, middle, and end (Cleary et al. 2012). The microanalytic questions can be open-ended or closed-ended with the result that the answers can have a quantitative or a qualitative format. The advantage of open questions is their nonsuggestive character. Nevertheless,

it is necessary to develop a formal set of criteria to categorize the assessed information. The largest difference of microanalysis to SRL questionnaires therefore is that microanalytic questions refer to specific steps of the learning process and are answered in close contact to the actual learning process whereas questionnaires are answered in a retrospective and more generalized way. Research has shown that microanalytic assessments have a good validity and reliability (Cleary et al. 2012; Cleary and Sandars 2011). The correlation between microanalytic assessments and SRL questionnaires is controversial as some investigations can find a moderate correlation (DiBenedetto and Zimmerman 2013), whereas other studies did not find any relation (Cleary et al. 2015). One reason could be a different grade of generalization and specificity in the two measures. However, microanalytic assessments seem to show an association with performance results and academic achievement (Cleary et al. 2015; DiBenedetto and Zimmerman 2010, 2013; Lau et al. 2015) and therefore can complement other SRL assessment methods (Cleary et al. 2012).

Whereas microanalytic assessments are task-specific but nevertheless self-reports, it is possible to precisely record learning process steps of an individual via trace data. Trace data can exist in the form of log files, event recordings, or navigation trails (Hadwin et al. 2007). Gathering trace data enables one to capture the dynamic character of SRL processes that occur in a sequence of events (Greene and Azevedo 2010). Moreover, trace data can be used to make conclusions about a learner's ability and strategic knowledge in the context of SRL and to connect this information with the results of other SRL assessment methods. An advantage of trace data is their precision and objectivity. For example, trace data in text learning can be gathered by using the method of analyzing a learner's text highlighting (Winne 2010). By highlighting, the learner expresses metacognitive control while cognitive events can be visualized.

Further methods for the investigation of SRL include learning diaries, interviews, observation procedures, or think-aloud protocols. In learning diaries, the gathering of information occurs continuously along a period of several weeks in specific learning processes. Although learning diaries can record variations in behavior sensitively (Schmitz and Wiese 2006), the learner's motivation may not be constant along the assessment period and could distort the data (Spörer and Brunstein 2006). Interviews can cover different aspects of SRL without information loss, but they show the abovementioned typical disadvantages of self-report measures, especially the influence of social desirability (Pospeschill 2010). Observation measures are suitable for testing younger children in SRL, but there are some limitations concerning the objectivity of judging the observed behavior (Whitebread et al. 2009). Think-aloud protocols help record a learner's unbiased thoughts while working on a specific task, but the verbalization itself could interfere with the learning process and task performance (Spörer and Brunstein 2006).

Indeed, several methods to assess SRL exist and they all show benefits and shortcomings. Therefore, several authors argue for the combination of self-reports with data that are more objective from think-aloud protocols, interviews, teacher ratings, observations, or trace data, as these combinations will ensure reliable results (Perry and Rahim 2011; Veenman 2011). The alignment of various assessment methods can compensate for several disadvantages of different instruments. To date, most studies did not use a multimethod approach for assessing SRL. Nevertheless, there already are some studies that have investigated the relationship between different SRL measures: For example, McCardle and Hadwin (2015) compared self-report questionnaire data with more fine-grained self-report data from weekly reflections of students. Using both data sources, they were able to extract SRL profiles and could cross-validate the questionnaire results with the results of weekly reflections. A study by Trevors et al. (2016) combined eye-tracking data, metacognitive

judgments, computer log files, and concurrent as well as retrospective verbal SRL reports. Using this methodology, the authors could gain new insight into the relation of epistemological beliefs and SRL. Cleary and Callan (2018) conducted a study with middle school math students that aimed at the multidimensional assessment of SRL using questionnaires and teacher ratings as broad-based measures, and microanalysis as well as behavioral traces as task-specific measures. They found moderate correlations within the assessment level (broad vs. specific) and no correlation between these levels. Only microanalytic monitoring scores as well as teacher ratings showed predictive validity for student achievement in a math test (Cleary and Callan 2018). Cleary et al. (2015) compared microanalytic results with questionnaire data in a sample of college students. They found high intercorrelations for the microanalytic measure and no correlation between microanalytic results and questionnaire data. Moreover, microanalytic results could be used as a predictor of future performance.

Purpose of the present study

As several authors argue for a multimethod assessment of SRL (e.g., Callan and Cleary 2017; Perry et al. 2002), the present study combined several assessment methods and aimed to investigate their interrelations as well as their relation to achievement. Using such a multimethod approach can help to gather information on congruencies and differences between assessment methods that are based on the same theoretical models but use different measurement levels. Moreover, it could be possible to find out which instruments are most appropriate for the assessment of different components of SRL, so that future studies could combine these instruments in accordance to the theoretical research questions. In accordance with theoretical conceptions of SRL (e.g., Boekaerts 1999) and in order to have a more specific picture of SRL related measures, we aimed to assess cognitive, metacognitive, and motivational components of SRL. According to Cleary and Callan (2018), we compared two aptitude measures, namely an SRL questionnaire and an SRL strategy knowledge test, that assessed the three components of SRL. Using two aptitude measures helped us to combine the strengths and weaknesses of both, while investigating if one method is able to outplay the other method concerning a specific component (e.g. if the motivational component can be assessed better through questionnaires or strategy knowledge tests). Moreover, we used two event measures, namely microanalytic assessments (open-ended and closed-ended) as well as trace data (behavioral traces). As metacognitive and motivational strategies are not observable, we used microanalytic questions for their assessment and used trace data to gain objective information on the use of cognitive strategies as these can be seen within the learning materials, e.g. making notes or underlining important text sections. We wanted to pursue the multimethod approach on event-level as well, but microanalytic assessments have the weaknesses of self-report measures. Therefore, we used the method of trace data to at least gain objective data on the use of cognitive strategies.

Due to possible problems with the autonomous learning setting and the self-guided way of acquiring relevant knowledge, SRL is especially important for college students. A valid assessment of SRL is necessary to develop useful training programs to foster this competence. This is why we analyzed the convergence between measures as well as the correlation to achievement within a sample of college students. Of special interest were the reliability and validity of the strategy knowledge test (as this was a newly developed instrument) and the microanalytic assessments and their potential advantages in contrast

to the questionnaire. We hypothesized that (1) measures within the same assessment levels (i.e. that both measures either refer to event-level or to aptitude-level, monomethod) show (higher) correlations than measures of different assessment levels (heteromethod). This hypothesis is underlined by the empirical findings of Rover et al. (2019), who state that different measures of SRL show higher convergence when the assessment is coarse grained and focuses on students' global use of self-regulated learning strategies. With regard to previous studies (e.g., Cleary et al. 2015), we moreover expect (2) microanalytic assessments to show the highest criterion validity for academic achievement.

Method

Sample

Our sample consisted of 71 college students of a southwestern German university (female = 77.5%). Mean age was 22 ($M = 21.93$, $SD = 2.87$, *Range* = 18–30). One person had to be excluded due to unusable data, and therefore the final sample consisted of 70 college students. Sixty-four students (91.4%) were psychology students and six students were student teachers. The average college entrance diploma GPA (Abitur) was $M = 1.76$ ($SD = 0.50$, *Range* = 1.1–3.3) and the average GPA in the subject of study was $M = 2.00$ ($SD = 0.53$, *Range* = 1.0–3.7). The students were in their third semester ($M = 3.46$, $SD = 2.28$, *Range* = 2–10). To investigate the abovementioned research questions, we conducted our tests in computerized and standardized settings. The data acquisition was completely anonymized, the participants had to sign an informed consent and had the possibility to withdraw their consent at any time of the study. Participants were rewarded with test person hours, candies, and participation in a shopping voucher lottery.

Procedure

Test sessions were realized in groups of up to a maximum of seven persons and were interchangeably guided by two trained psychology research assistants. All questionnaires were filled out within a computerized survey program. At the beginning of the test, demographic data were assessed (age, gender, field of study, semester of study, college entrance diploma GPA). Afterwards, an SRL strategy knowledge test was presented. As we needed a specific learning task for the event measures (microanalyses and trace data), students had to read and prepare a text on catastrophe management (a subject where we expected low pretest knowledge) with the aim of answering a corresponding knowledge test. In this context, the microanalytic SRL assessment tool was implemented. Lastly, an SRL questionnaire was administered so that self-referring and suggestive effects on the other SRL measures could be avoided.

Instruments

Aptitude measure I: strategy knowledge test on SRL

First, participants were asked to answer a strategy knowledge test that was constructed based on Zimmerman's (2000) and Boekaerts' (1999) models of SRL (cognitive,

metacognitive, and motivational components within planning, performance, and reflection phase of learning). In this context, an academic learning scenario on the preparation of an oral presentation on the subject of educational psychology was presented. Within a short description, the problems of a female student (for female participants) or a male student (for male participants) were expounded. In general, the character's interest in this subject was low. Specifically, the scenario was divided into the three phases of SRL (forethought, performance, and self-reflection) and for each phase, theoretically useful and not useful metacognitive and motivational strategies were presented. For the performance phase, cognitive strategies were presented as well. As an example, we described the motivational problem in the planning phase: "Phil has to prepare a graded presentation within four weeks in cooperation with one of his fellow students. The presentation should deal with an issue of educational psychology but Phil is totally not interested in this subject. Therefore, he accepts the issue proposal of his fellow student. As how useful do you rate the following strategies for preparing the presentation? ". The participants had to rate the usefulness of learning strategies with regard to the situation described in the scenario on a four-point rating scale (1 = *not useful at all*, 4 = *very useful*), whereby always two statements represented pairs of a useful and a less useful strategy referring to the same component. At the beginning of the test, there was an explicit hint that the task did not consist of choosing a strategy that represents someone's own typical practice, but in rating the general usefulness of the strategy (for an example, see Table 2). The categorization into useful and not useful strategies was based on theoretical assumptions as well as expert evaluations.

To calculate a score for each participant, the relative judgment of superiority respective of inferiority of strategies was applied (referring to Artelt et al. 2009). In other words, it was scored if the participants recognized the more useful strategy within a comparison with a less useful strategy (within a phase and a component). Thus, the strategy knowledge of SRL was defined by the score participants obtained in 25 paired comparisons. If participants rated the less useful strategy as equally or more useful as the useful strategy (no difference or a negative difference), 0 points were given. If participants rated the useful strategy one point higher than the less useful strategy (a difference of 1), 1 point was

Table 2 Item examples, number of comparisons, item difficulties, and discriminatory power of the strategy knowledge test Items

Scale	Number of comparisons	Comparison example	*Range* P_i	*Range* r_{it}
Metacognition	9	He should ensure not to lose sight of his goal (+) While compiling the presentation, he should not think about what he is doing. This would merely confuse him (−)	28–89	0.19–0.55
Motivation	7	He should look for ways to make the topic of the presentation as interesting as possible for himself (+) He should accept his fellow student's sparsely interesting topic proposal so that he won't have to think about it anymore (−)	59–94	0.34–0.74
Cognition	9	He should critically examine the author's depiction of results (+) He should read the text five times (−)	58–74	0.33–0.66

(+) useful strategy, (−) less useful strategy, P_i = item difficulty, r_{it} = discriminatory power

given, and if participants rated the useful strategy 2 or 3 points higher than the less useful strategy (a difference of 2 or 3), 2 points were given. Thus, our scoring followed the scoring of previous studies using strategy knowledge tests (e.g. Händel et al. 2013; Karlen 2017), but we obtained to gain a more fine-grained rating and therefore used graduations in scores according to how close the rating was to the expert rating. The maximum attainable sum score therefore was 50 (25 paired comparisons) and the average sum score was $M = 33.93$ ($SD = 7.08$). For our analyses, we calculated mean scores for the three components of motivation ($M = 1.63$, $SD = 0.37$), metacognition ($M = 1.21$, $SD = 0.38$), and cognition ($M = 1.28$, $SD = 0.50$). Examples of the 25 comparisons, number of comparisons, difficulties, and discriminatory power of comparisons are shown in Table 2. The reliability index measured by Cronbach`s alpha was 0.71 for the metacognition scale, 0.76 for the motivation scale, and 0.84 for the cognition scale.

Aptitude measure II: SRL questionnaire

To assess students' SRL via questionnaire, an instrument of Dörrenbächer and Perels (2016) was used, supplemented by items of Schwinger et al. (2007) for the motivational component. The items were slightly adapted to make them comparable to the strategies used in the strategy knowledge test. The measure consisted of items on elaborative learning strategies for the cognitive component (4 items, $\alpha = 0.66$), items on goal setting, time planning, self-monitoring, self-evaluation, and self-reaction for the metacognitive component (19 items, $\alpha = 0.81$), and items on attention focusing, goal orientation, procrastination (reverse coded), intrinsic motivation, and environment control for the motivational component (20 items, $\alpha = 0.86$). The items were rated on a four-point rating scale (1 = *not true at all*, 4 = *totally true*). For further information about the questionnaire, please see Dörrenbächer and Perels (2016). To prevent possible suggestive effects on the other SRL instruments, the questionnaire was applied last in our test setting.

Event measure I: microanalytic assessment

The microanalytic assessment instrument was constructed based on Zimmerman's (2000) and Boekaerts' (1999) models of SRL. It consisted of some new formulated items and some items based on former studies by Artino et al. (2014), Callan (2014), Cleary and Callan (2018), Cleary and Zimmerman (2004), and DiBenedetto and Zimmerman (2010). The microanalytic assessment aimed at evaluating the usage of SRL strategies referring to a specific learning situation. Within a timeframe of 10 min, the participants had the task to learn a text on catastrophe management (Grün and Schönenberger 2018) with the goal of preparing for a knowledge test afterwards. Microanalytic questions referring to the forethought phase were asked after the participants had a short overview of the learning task. After five minutes of learning the text, microanalytic questions referring to the performance phase were presented. After accomplishing the learning task and the associated knowledge test, microanalytic questions referring to the self-reflection phase were asked. The microanalytic questions addressed the two SRL components metacognition and motivation (as metacognitive and motivational strategies are not observable, we used microanalytic questions for their assessment and used trace data to gain information on the use of cognitive strategies as these can be seen within the learning materials, e.g. making notes or underlining important text sections). Questions were either open-ended or closed-ended with a four-step rating scale. Table 3 shows the closed-ended microanalytic forethought

Table 3 Microanalytic closed-ended items with item difficulty indices and discriminatory power

Scale	Item	P_i	r_{it}
Metacognition (planning and self-monitoring)	I've planned to use special strategies for task performance	0.61	0.62
	I have a plan for my procedure at task performance	0.64	0.53
	I pay attention not to lose sight of my goal	0.43	0.56
	I constantly review my procedure while task performance	0.60	0.34
Motivation (self-efficacy)	I am sure that I can master this task	0.71	0.51
	I will master this task successfully	0.53	0.51

P_i = item difficulty, r_{it} = discriminatory power

phase items as examples. The open answers were categorized and coded by two independent, trained raters to prepare the data for quantitative analysis. The coding scheme for self-motivation was constructed in a theory-based way (Schiefele and Schaffner 2015) and the coding scheme for task-preparation was theoretically developed in keeping with learner`s answers. A coding scheme example for the open-ended questions is displayed in Table 4.

Due to reasons of parsimony, for the following analyses, we chose closed-ended questions on self-efficacy (motivational component, 2 items, $\alpha = 0.66$) and planning and self-monitoring (metacognitive component, 4 items, $\alpha = 0.71$, see Table 3) as well as open-ended questions on self-motivation ("How do you motivate yourself for this task?", ICC = 0.79) and strategic planning ("What are you thinking about while you are preparing for this task?", ICC = 0.74).

Event measure II: trace data

As the used microanalytic questions focused on motivational and metacognitive strategy usage, we used trace data to assess the usage of cognitive strategies. Trace data contain no self-report as they can be found within the learning material and therefore are not prone to subjective distortion. While working on the text on catastrophe management, participants could use pens and highlighters by choice but were not instructed to do so. It was then possible to analyze a learner`s text highlighting and notes as trace data. Thereby, we distinguished between highlighted key words and highlighted auxiliary words to draw conclusions about a learner`s cognitive judgments of relevant text information (Leutner et al. 2007). In this study, key words are nouns and explaining verbs and adverbs, for example, "systematic observation" or "disastrous event". By contrast, auxiliary words are helpful to understand grammatical constructions, but not to apprehend the text content. Examples are pronouns, conjunctions, or modal verbs. Moreover, notes within the text were coded by two independent raters with regard to the representation of cognitive learning strategies (Wild and Schiefele 1994; ICC = 0.91). The coding scheme is displayed in Table 5.

Academic achievement

As a criterion variable, we used GPA of college entrance diploma (Abitur) as this is a central exam in Germany with high relevance for educational careers. This score is reverse coded (i.e., higher scores indicate lower school performance). Moreover, we planned to use the results of the knowledge test on catastrophe management, as this task had to be

Table 4 Coding scheme for open-ended microanalytic questions

Question	Category	Definition	Coding score	Answer example
How do you motivate yourself for this task? (self-motivation)	Answer without expressiveness	Answer does not contain the asked information	−99	"I need no motivation." "No statement."
	Missing/vague motivation	No motivation, discernible/motivation, unratable	0	"Few." "I stay relaxed." "Mediocre."
	Extrinsic motivation	Extrinsic motivation is the intention of accomplishing a learning process due to its consequences (Schiefele 1996)	1	"Results will be controlled." "I don`t want to do badly." "Test person hours are my motivation."
	Intrinsic motivation	Intrinsic motivation is defined as the desire to accomplish an activity because it is accompanied by positive experiences (Schiefele 1996)	2	"I think this theme is interesting." "I would like to support the scientific research." "I want to test my own abilities." "Motivation to learn something new."
	Extrinsic and intrinsic motivation	Coexistence of extrinsic and intrinsic motivation as set out above	3	"Sounds interesting and I will get test person hours in return."

Table 4 (continued)

Question	Category	Definition	Coding score	Answer example
What are you thinking about while you are preparing for this task? (strategic planning)	Missing statement Confounder Task-irrelevant factors	Answer is missing/contains no information Thinking about disturbing conditions of task performance, does not contribute to goal attainment Conceptual engagement in issues outside of the learning situation	0	"Nothing special." "About the pressure of time." "About insufficient concentration." "How difficult the questions may be."
	Task-related factors	Principle engagement in the task	1	"About the text."
	Content-orientation	Thoughts about the theme content or activation of prior knowledge	2	"What could be the most important information." "What range of subjects is there in the context of disaster management? What do I already know? What am I interested in?"
	Task-orientation	Thoughts about working on the task, what is judged as target-aimed	2	"Time management, text structuring, finding main points." "What details could be important and what kind of questions will be asked." "Strategy for working on the task."
	Task- and content-orientation	Topical as well as strategic engagement with task	3	"I try to find coherences to facts I have learned or experienced before. For example, I connect text content and pictures of media reports on catastrophes in the past."

A code of −99 was given, if the answer could not be used for analyses due to missing information

Table 5 Coding scheme of cognitive learning strategies within the learner`s notes

Category	Definition	Coding score	Examples
No notes		0	
No strategy obvious	Notes exist in form of incoherent or not assignable keywords or the extent is slight	1	
Notes for organization: formal structuring of text content	Notes are servant for building an overview of text structure	2	Section headings
Notes for organization: content summary/questions at the text	Notes are servant for providing a content-related overview of the text or questions are formulated to support the learning process	3	"Catastrophe management Methods for avoidance, rescue etc Through analysis and observation of catastrophes" "What is the catastrophic event and process?"
Notes for elaboration/learner`s own thoughts	Notes testify a content engagement in text content and beyond the text content	4	Drawing a diagram about precaution in dependence of frequency and severity of incoming catastrophes, finding own examples
Notes for organization and elaboration	Notes provide a formal or content-wise overview of the text and beyond the text	5	

completed during the test session. The test is comprised of eight multiple-choice items on the content of the text. Each multiple choice item had four answer possibilities that could be right or wrong, each possibility that was answered correctly was scored with one point. Therefore, the maximum test score was 32. The mean score of 28.96 ($SD = 1.76$) shows strong ceiling effects for the test. Only 5 out of 32 answer possibilities showed a difficulty below 0.80. Due to this low variance, we decided not to use the test results as a valid marker for academic achievement.

Results

In line with previous studies that assessed SRL in a multidimensional way (e.g., Callan and Cleary 2017; Cleary et al. 2015), we analyzed correlations within (monomethod) as well as between (heteromethod) the general and specific assessment levels (H1). In doing so, we focused not only on SRL in general, but on the correlations of the three components to gain new insight into their relations. Moreover, we examined the correlations of both instruments with achievement, in order to examine if microanalytic assessments show the highest criterion validity for academic achievement (H2).

Heterotrait-monomethod correlations

The results referring to the correlations between the three different components, within one assessment level, can be found in Table 6. Although moderate correlations between metacognition and motivation as well as cognition in the strategy knowledge test are evident, there is no correlation between cognition and motivation in this test. With regard to the questionnaire, there is a rather high correlation between metacognition and motivation, but no correlation between cognition and motivation as well as metacognition. Referring to the microanalytic assessment, there are moderate correlations for the open-ended as well as for the closed-ended questions. Concerning trace data, only a high correlation between the number of highlighted key words and the number of highlighted auxiliary words can be found.

Table 6 Correlations between the three different components within one assessment level

Strategy knowledge test	MOT – COG	– 0.19
	MOT – MCOG	0.28*
	COG – MCOG	0.34**
Questionnaire	MOT – COG	– 0.08
	MOT – MCOG	0.54**
	COG – MCOG	– 0.06
Microanalytic assessment–closed-ended	MOT – MCOG	0.26*
Microanalytic assessment–open-ended	MOT – MCOG	0.27*
Trace data	CLS – KW	0.10
	CLS – AW	0.07
	KW – AW	0.72**

MOT motivation; *MCOG* metacognition; *COG* cognition; *CLS* cognitive learning strategies; *KW* key words; *AW* auxiliary words

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

Monotrait-heteromethod correlations

To investigate the relation between one component assessed by different methods, correlations were computed. Table 7 shows the results for the correlations of the same component across different assessment methods. No correlations were found except for the motivational component assessed by questionnaire and closed-ended microanalytic questions.

Correlations with academic achievement

As we wanted to analyze the criterion validity of the three components and the different assessment methods, we computed correlations to GPAs of the college entrance diploma. The results are shown in Table 8. We found several correlations of moderate size (i.e., for the cognitive component assessed via strategy knowledge test, for the motivational component assessed via questionnaire, and the metacognitive component assessed via one open-ended microanalytic question). As GPA in Germany is reverse coded, negative correlations speak in favor of the criterion validity of these measures. Moreover, we found a positive correlation between the number of highlighted auxiliary words and GPA.

Table 7 Results for the correlations of the same component across different assessment methods

Motivation	SKT – Q	0.09
	SKT – MI_C	0.05
	SKT – MI_O	– 0.02
	Q – MI_C	0.29*
	Q – MI_O	0.00
	MI_C – MI_O	0.14
Metacognition	SKT – Q	0.22
	SKT – MI_C	0.17
	SKT – MI_O	0.03
	Q – MI_C	0.21
	Q – MI_O	0.10
	MI_C – MI_O	0.16
Cognition	SKT – Q	0.05
	SKT – TD_CLS	0.12
	SKT – TD_KW	– 0.04
	SKT – TD_AW	– 0.13
	Q – TD_CLS	0.15
	Q – TD_KW	– 0.05
	Q – TD_AW	– 0.06

SKT strategy knowledge test; *Q* questionnaire; *MI_C* microanalysis_closed-ended questions; *MI_O* microanalysis_open-ended questions; *TD_CLS* trace data_cognitive learning strategies; *TD_KW* trace data_key words; *TD_AW* trace data_auxiliary words

$^{*}p < 0.05$, $^{**}p < 0.01$

Table 8 Correlations of the three components and different assessment methods to GPA of college entrance diploma

Measure	GPA
Strategy knowledge test—motivation	0.00
Strategy knowledge test—cognition	−0.33**
Strategy knowledge test—metacognition	−0.15
Questionnaire—motivation	−0.24*
Questionnaire—cognition	−0.02
Questionnaire—metacognition	−0.20
Microanalysis—motivation—closed-ended questions	−0.19
Microanalysis—metacognition—closed-ended questions	−0.07
Microanalysis—motivation—open-ended questions	−0.10
Microanalysis metacognition—open-ended questions	−0.26*
Trace data—cognitive learning strategies	−0.08
Trace data—key words	−0.08
Trace data—auxiliary words	0.29*

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

Discussion

The purpose of the present study was to perform a multimethod assessment of SRL in a sample of college students to investigate the interrelations of different assessment methods as well as their correlation to academic achievement. In general, we found moderate correlations within one assessment level (aptitude or event), but only for some components. Across the assessment levels, few significant correlations were found. With regard to their correlation to GPA, we found differing assessment methods for every component to be related to academic achievement (cognitive component measured by strategy knowledge test, motivational component measured by questionnaire, and metacognitive component measured by open-ended microanalytic question). In the following, we will discuss these results in depth.

Instruments used for the assessment of SRL

Following the multimethod approach, we used two instruments for each assessment level. To achieve this aim, we had to develop a strategy knowledge test and a microanalytic approach based on existing literature. As they were not used previously, we will discuss each instrument briefly. Regarding the psychometric information, the reliability for both instruments is satisfying. To discriminate best between participants, items should not be too easy nor too difficult. For the microanalysis, item difficulties vary between 43 and 64%, which is acceptable due to the convention of perfect item difficulties ranging about 50% (Schmidt-Atzert and Amelang 2012). The strategy knowledge test reveals a greater variance concerning the difficulties of comparisons (28%–94%), which means that some comparisons were too easy and others too difficult to solve. In accordance with convention (Moosbrugger and Kelava 2012), discriminatory power for the microanalytic questions was between 0.40 and 0.70. In contrast, results of the strategy knowledge test were less consistent and therefore revealed a starting point for further development.

Heterotrait-monomethod relations

Concerning the results of interrelation of different assessment measures for the three components of SRL (motivation, metacognition, and cognition), our results partly accord with those of previous studies. Within one assessment level (aptitude or event), we found moderate correlations, comparable with Callan and Cleary (2017), but only for some components. Within the aptitude level, we found significant correlation between motivation and metacognition and between cognition and metacognition. Obviously, there are similar patterns for the aptitude measures and their intercorrelations of the components. As the cognitive component assessed by the questionnaire does not show any meaningful relation to the other components, it could be hypothesized that the self-assessment of cognitive strategy use is based on a different mechanism than the self-assessment of motivational or metacognitive strategy use. As cognitive strategies are rather concrete ("I create tables or diagrams to organize the meaning of the text"), they probably can be judged more validly than metacognitive or motivational strategies that are not as easy to observe ("After finishing studying, I think about what I could have done better"). Regarding the correlations for the components within the event measures, significant correlations were found for the closed-ended microanalytic questions as well as for the open-ended questions concerning motivation and metacognition. This is in line with results of Cleary et al. (2015). Cognition was assessed by trace data, but there were no correlations between the applied cognitive learning strategies and the number of highlighted key words or auxiliary words.

Monotrait-heteromethod relations

Regarding the interrelations within one component across several assessment methods, few significant correlations could be found, similar to Cleary et al. (2015). This means that neither the assessed aptitude measures nor the assessed event measures seem to cover the same constructs. The only significant correlation was found for motivation in the questionnaire and the closed-ended microanalytic question. This correlation probably results from the fact that both questionnaire and closed-ended microanalytic questions used four-point rating scales for the assessment.

Our findings suggest that there is more than the discrimination between aptitude and event measures. Few significant correlations were found across different assessment methods and correlation patterns of components differed clearly from one assessment method to another. This can be due to the fact that the SRL questionnaire, for example, did not only assess declarative knowledge of learning strategies but also their application. Equivalently, the microanalytic assessment and the trace data comprised different components of SRL, and therefore it is not surprising that no significant relations can be found. One can argue that there has to be a relation between knowledge and application of strategies, but in fact, the application is always dependent on situational factors (Artelt 2000; Pressley et al. 1987), which can explain the missing correlation.

Correlation to achievement

With regard to the relationship of the different assessment methods with achievement, we conducted correlations with the GPAs of college entrance diplomas for every

component and expected significant correlations (Artelt et al. 2012; Broadbent 2017; Cleary and Callan 2018; DiBenedetto and Zimmerman 2010). Interestingly, for every component, we found another assessment method to be correlated with academic achievement: The cognitive component showed a significant relation if measured by the strategy knowledge test, the motivational component if measured by questionnaire, and the metacognitive component if measured by an open-ended microanalytic question. Other components of the different assessment methods did not show any significant relation to GPA.

It could be hypothesized that the three components of SRL bring different characteristics that are then best measured with different assessment methods. The cognitive component is strongly influenced by the learner's knowledge on learning strategies and their appropriate use (Boekaerts 1999). Therefore, it seems valuable to assess this component with a strategy knowledge test. The motivational component represents strategies to motivate oneself but also beliefs and attitude (e.g., self-efficacy and goal orientation; Pintrich 2004; Wolters 2003). As beliefs can be assessed validly via questionnaires (Pintrich et al. 1993), it seems justifiable to measure this component using questionnaires. The metacognitive component represents second-order thinking and the regulation of strategy use. As this component is highly situation-specific and depends on the requirements of a concrete learning process, it seems likely that such processes can be measured best using event measures, such as microanalytic assessments. Concluding, it seems justifiable to argue for a multimethod approach in future studies on SRL. Using different methods for different components of SRL could be more economic than measuring each component with each assessment method. In line with this, Rovers et al. (2019) found the granularity of assessment to be important when comparing SRL measures that focus on different assessment levels. It seems that different measures show higher convergence and that students are better in self-reporting on SRL strategy usage if global use of SRL strategies is assessed, while the convergence is low if the use of concrete SRL strategies is in focus.

Limitations

In the following, we would like to discuss some aspects that limit our results and their generalization and would have to be optimized for future research. First, due to problems in the acquisition of participants for the elaborate study our sample was based on only 70 undergraduates and was therefore smaller than planned. A priori power analysis using G*Power (Faul et al. 2007) for two-tailed correlations with a moderate size ($\rho = 0.30$), an alpha level of 0.05 and a power of 0.80 resulted in a sample size of $n = 84$. As the testing procedure lasted for two hours at least, it was not possible to meet this sample size, as most students do not have this amount of free time between their courses. Moreover, we had a time frame for the study and did not achieve the sample size by the end of this time. Consequently, it is possible that we were not able to detect all significant effects in the data. Moreover, the sample was extremely selective, as it was comprised of 90% of students in psychology. We selected only students who had not attended a lecture in pedagogical psychology, but we cannot be entirely certain about their prior knowledge regarding SRL.

Furthermore, a limitation of this study exists in the test session itself as it was realized in groups of up to seven students for reasons of test economy. It can be assumed that with an increasing number of persons in the test room, disturbing effects that are attributable to background noise and visual distraction might emerge. Questionnaires were completed

within a computerized survey program, but the texts for the text-learning task on catastrophe management were given in printed versions to allow students to highlight key words or make notes. Possibly, students who saw their neighbor highlighting words and making notes were likely to do the same, although they would not have had the idea themselves. Moreover, students having questions might have been too shy to ask them in front of the group. Additionally, for reasons of standardization, test persons were required to spend at least 10 min in their learning process and thus were restricted in their time planning.

Also, it could be possible that the presentation of microanalytic questions about the performance phase interrupted the learning process, which could make the results less comparable to usual learning processes. The verbalization of the learning process could also cause interferences with the learning process itself by inducing a deceleration (Veenman et al. 2003). Another possibility is that the microanalytic questions themselves could encourage self-regulated behavior, comparable to learning diaries (e.g., Klug et al. 2011). Beyond that, different question formats possess different advantages and disadvantages: Closed-ended questions show a clear structure and enable an economic and comparable assessment, but they provide superficial information and can cause approval or refusal tendencies. Answering open-ended questions in return provides detailed and reflected information and suggestive effects can be minimized. Nevertheless, they also depend on motivation effects (Holland and Christian 2009), verbal abilities (Züll 2014), and an information loss emerges through the process of categorizing and coding (Kempf 2010). A limitation of the microanalytic assessment instrument is the absence of a cognitive scale similar to the metacognitive and motivational scale due to the intention to capture the cognitive learning strategies by trace data. Moreover, the cognitive scale was only assessed by four items in the questionnaire. Therefore, this component was measured only sparsely and the values maybe are not as informative as the one of the motivational and metacognitive scale. In further studies, the cognitive component should be measured more comprehensively. In general, it is problematic that we could not realize measurements of all components with all assessment methods.

Another limitation concerns the assessment of academic achievement. For correlations, we used college entrance diploma GPA, which is a highly global measure as it represents the mean of several learning results in high school. This is why it seems unlikely that it would reveal correlations with event measures, as they are highly situation-specific. Therefore, we implemented a text knowledge test, but unfortunately, we found ceiling effects because the test was too easy and consequently we were not able to conduct analyses with this test. Probably, we would have found meaningful correlations with this test and at least the event measures.

Implications for research and practice

Multimethod assessment of SRL in college students seems to be a promising research approach. It appears that different assessment methods are necessary to optimally capture different SRL components since our study gives first hints that cognition can validly be gathered by the strategy knowledge test, metacognition can validly be assessed by the SRL microanalysis, and motivation can validly be captured by the SRL questionnaire. Therefore, it could be useful to develop a test battery that combines different assessment methods for the different SRL components. Of course, a replication of these results is required. Additionally, further factors for validity examination could be incorporated in replication studies

(e.g., measuring all components of SRL with all assessment methods) and single-person tests can be taken into consideration. Alternatively, it could be asked, in what aspects the implemented learning process resembles a learner's usual learning process. Future investigations could explore the reasons learners use some learning strategies and avoid others. Based on the production deficit (Hasselhorn 1996; Spörer and Brunstein 2006) and the distinction between the knowledge of learning strategies and the competence to apply these strategies, it is of special interest if a decision against a strategy rests upon a lack of motivation, missing metacognitive competencies, or an overestimation of someone's own abilities. Hereby, special needs for SRL training could be found that enable fostering student`s SRL in designated facets (Händel et al. 2013) so that one of the main goals of research in SRL is pursued.

With regard to educational practice, SRL test batteries could be used in interventional studies to gain a closer look on benefits resulting from specific interventions on SRL strategies. In addition, when planning and creating interventions such as strategy trainings for schools and learners, a valid assessment and diagnosis of the status quo is necessary. Using a test battery that combines assessment methods for different components of SRL could help to gain a broader view on SRL skills of learners and to create more tailored and individualized interventions instead of general interventions that are the same for all learners.

Funding Open Access funding enabled and organized by Projekt DEAL.

Compliance with ethical standards

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

Ethical approval All procedures followed were in accordance with the ethical standards of the responsible committee (The Ethics Committee of the Faculty for Empirical Human Sciences and Economical Sciences, Saarland University). Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Artelt, C. (2000). *Strategisches Lernen [Strategic learning]*. Münster: Waxmann.

Artelt, C., Beinicke, A., Schlagmüller, M., & Schneider, W. (2009). Diagnose von Strategiewissen beim Textverstehen [Diagnosis of strategic knowledge in text comprehension]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 41*(2), 96–103. https://doi.org/10.1026/0049-8637.41.2.96.

Artelt, C., Naumann, J., & Schneider, W. (2010). Lesemotivation und Lernstrategien [Reading motivation and learning strategies]. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, et al. (Eds.), *Pisa 2009, Bilanz nach einem Jahrzehnt [Pisa 2009, balance after one decade]* (pp. 73–112). München: Waxmann.

Artelt, C., Neuenhaus, N., Lingel, K., & Schneider, W. (2012). Entwicklung und wechselseitige Effekte von metakognitiven und bereichsspezifischen Wissenskomponenten in der Sekundarstufe [Development

and reciprocal effects of metacognitive and domain-specific knowledge components in secondary education]. *Psychologische Rundschau, 63,* 18–25. https://doi.org/10.1026/0033-3042/a000106.

Artino, A. R., Jr., Cleary, T. J., Dong, T., Hemmer, P. A., & Durning, S. J. (2014). Exploring clinical reasoning in novices: A self-regulated learning microanalytic assessment approach. *Medical Education, 48*(3), 280–291. https://doi.org/10.1111/medu.12303.

Azevedo, R., Johnson, A., Chauncey, A., & Graesser, A. (2011). Use of hypermedia to assess and convey self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 102–121). New York: Routledge.

Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist, 45*(4), 210–223.

Bail, F. T., Zhang, S., & Tachiyama, G. T. (2008). Effects of a self-regulated learning course on the academic performance and graduation rate of college students in an academic support program. *Journal of College Reading and Learning, 39*(1), 54–73. https://doi.org/10.1080/10790195.2008.10850312.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bellhäuser, H., Lösch, T., Winter, C., & Schmitz, B. (2016). Applying a web-based training to foster self-regulated learning—Effects of an intervention for large numbers of participants. *The Internet and Higher Education, 31,* 87–100. https://doi.org/10.1016/j.iheduc.2016.07.002.

Bembenutty, H. (2011). New directions for self-regulation of learning in postsecondary education. *New Directions for Teaching and Learning, 126,* 117–124. https://doi.org/10.1002/tl.450.

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research, 31*(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2.

Broadbent, J. (2017). Comparing online and blended learner's self-regulated learning strategies and academic performance. *The Internet and Higher Education, 33,* 24–32. https://doi.org/10.1016/j.iheduc.2017.01.004.

Callan, G. L. (2014). *Self-regulated learning (SRL) microanalysis for mathematical problem solving: A comparison of a SRL event measure, questionnaires, and a teacher rating scale.* Dissertation: University of Wisconsin-Milwaukee.

Callan, G. L., & Cleary, T. J. (2017). Multidimensional assessment of self-regulated learning with middle school math students. *School Psychology Quarterly, 33*(1), 103–111. https://doi.org/10.1037/spq0000198.

Cleary, T. J. (2011). Emergence of self-regulated learning microanalysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 329–345). New York: Routledge.

Cleary, T. J., & Callan, G. L. (2018). Assessing self-regulated learning using microanalytic methods. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 338–351). New York: Routledge.

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.

Cleary, T. J., Callan, G. L., & Zimmerman, B. J. (2012). Assessing self-regulation as a cyclical, context-specific phenomenon: Overview and analysis of SRL microanalytic protocols. *Education Research International*. https://doi.org/10.1155/2012/428639.

Cleary, T. J., & Sandars, J. (2011). Assessing self-regulatory processes during clinical skill performance: A pilot study. *Medical Teacher, 33*(7), 368–374. https://doi.org/10.3109/0142159X.2011.577464.

Cleary, T. J., & Zimmerman, B. J. (2004). Self-regulation empowerment program: A school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools, 41*(5), 537–550. https://doi.org/10.1002/pits.10177.

Cohen, M. T. (2012). The importance of self-regulation for college student learning. *College Student Journal, 46*(4), 892–902.

DiBenedetto, M. K., & Bembenutty, H. (2013). Within the pipeline: Self-regulated learning, self-efficacy, and socialization among college students in science courses. *Learningand Individual Differences, 23,* 218–224. https://doi.org/10.1016/j.lindif.2012.09.015.

DiBenedetto, M. K., & Zimmerman, B. J. (2010). Differences in self-regulatory processes among students studying science: A microanalytic investigation. *The International Journal of Educational and Psychological Assessment, 5*(1), 2–24.

DiBenedetto, M. K., & Zimmerman, B. J. (2013). Construct and predictive validity of microanalytic measures of students' self-regulation of science learning. *Learning and Individual Differences, 26,* 30–41. https://doi.org/10.1016/j.lindif.2013.04.004.

Dörrenbächer, L., & Perels, F. (2016). More is More? Evaluation of interventions to foster self-regulated learning in college. *International Journal of Educational Research, 78,* 50–65. https://doi.org/10.1016/j.ijer.2016.05.010.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement. Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Gläser-Zikuda, M., Voigt, C., & Rohde, J. (2010). Förderung selbstregulierten Lernens bei Lehramtsstudierenden durch portfoliobasierte Selbstreflexion [Fostering teacher trainee`s self-regulated learning via portfolio-based self-reflection]. In M. Gläser-Zikuda (Ed.), *Lerntagebuch und Portfolio aus empirischer Sicht [Learning diary and portfolio from empirical view]* (pp. 142–163). Landau: Empirische Pädagogik.

Grassinger, R. (2011). Selbstregulation beim Wechseln der Lernumwelt [Self-regulation on changing the learning environment]. In M. Dresel & L. Lämmle (Eds.), *Motivation, Selbstregulation und Leistungsexzellenz [Motivation, self-regulation and achievement excellency]* (pp. 179–197). Münster: LIT-Verlag.

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*(4), 203–209.

Grün, O., & Schönenberger, L. (2018). Betriebswirtschaftliches Katastrophenmanagement [Economical disaster management]. In M. Sulzberger & R. J. Zaugg (Eds.), *ManagementWissen [Management-Knowledge]* (pp. 237–243). Wiesbaden: Springer Gabler.

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. https://doi.org/10.1007/s11409-007-9016-7.

Händel, M., Artelt, C., & Weinert, S. (2013). Assessing metacognitive knowledge: Development and evaluation of a test instrument / Erfassung metakognitiven Wissens: Entwicklung und Evaluation eines Testinstruments. *Journal for Educational Research Online, 5*(2), 162–188.

Hasselhorn, M. (1996). *Kategoriales Organisieren bei Kindern. Zur Entwicklung einer Gedächtnisstrategie* [Children`s categorial organization. On the development of a memory strategy]. Göttingen: Hogrefe.

Holland, J. L., & Christian, L. M. (2009). The influence of topic interest and interactive probing on responses to open-ended questions in web surveys. *Social Science Computer Review, 27*(2), 196–212.

Karlen, Y. (2017). The development of a new instrument to assess metacognitive strategy knowledge about academic writing and its relation to self-regulated writing and writing performance. *Journal of Writing Research, 9*(1), 61–68. https://doi.org/10.17239/jowr-2017.09.01.03.

Kempf, W. (2010). *Quantifizierung qualitativer Daten [Quantification of qualitative data]*. Berlin: Regener.

Kitsantas, A. (2002). Test preparation and performance: A self-regulatory analysis. *The Journal of Experimental Education, 70*(2), 101–113. https://doi.org/10.1080/00220970209599501.

Kitsantas, A., Winsler, A., & Huie, F. (2008). Self-regulation and ability predictors of academic success during college. A predictive validity study. *Journal of Advanced Academics, 20*(1), 42–68. https://doi.org/10.4219/jaa-2008-867.

Klug, J., Ogrin, S., Keller, S., Ihringer, A., & Schmitz, B. (2011). A plea for self-regulated learning as a process: Modelling, measuring and intervening. *Psychological Test and Assessment Modeling, 53*(1), 51–72.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*(5), 178–181.

Lau, C., Kitsantas, A., & Miller, A. (2015). Using microanalysis to examine how elementary students self-regulate in math: A case study. *Procedia-Social and Behavioral Sciences, 174,* 2226–2233. https://doi.org/10.1016/j.sbspro.2015.01.879.

Leopold, C., & Leutner, D. (2002). Der Einsatz von Lernstrategien in einer konkreten Lernsituation bei Schülern unterschiedlicher Jahrgangsstufen [The usage of learning strategies in a concrete learning situation by pupils of different age-group levels]. *Zeitschrift für Pädagogik Beiheft, 45,* 240–258.

Leutner, D., Leopold, C., & Elzen-Rump, D. (2007). Self-regulated learning with a text-highlighting strategy: A training experiment. *Zeitschrift für Psychologie/Journal of Psychology, 215*(3), 174.

Lingel, K., Götz, L., Artelt, C. & Schneider, W. (2014). *MAESTRA 5–6+ - Mathematisches Strategiewissen für fünfte und sechste Klassen* [MAESTRA 5–6+ - Mathematical strategy knowledge for fifth and sixth grades]. PSYNDEX Tests Info.

Maag Merki, K., Ramseier, E., & Karlen, Y. (2013). Reliability and validity analyses of a newly developed test to assess learning strategy knowledge. *Journal of Cognitive Education and Psychology, 12*(3), 391–408.

McCardle, L., & Hadwin, A. F. (2015). Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. *Metacognition and Learning, 10*(1), 43–75.

Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion [Test theory and questionnaire construction]*. Berlin: Springer.

Neuenhaus, N., Artelt, C., & Schneider, W. (2017). Lernstrategiewissen im Bereich Englisch. Entwicklung und erste Validierung eines Tests für Schülerinnen und Schüler der frühen Sekundarstufe [Learning strategy knowledge in the subject of English. Development and first validation of a test for pupils of secondary education]. *Diagnostica, 63*(2), 135–147.

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research, 60*(6), 723–735.

Park, C. L., Edmondson, D., & Lee, J. (2012). Development of self-regulation abilities as predictors of psychological adjustment across the first year of college. *Journal of Adult Development, 19*(1), 40–49. https://doi.org/10.1007/s10804-011-9133-z.

Perels, F., Dignath, C., & Schmitz, B. (2009). Is it possible to improve mathematical achievement by means of self-regulation strategies? Evaluation of an intervention in regular math classes. *European Journal of Psychology of Education, 24*(1), 17–32. https://doi.org/10.1007/BF03173472.

Peverly, S. T., Brobst, K. E., Graham, M., & Shaw, R. (2003). College adults are not good at self-regulation: A study on the relationship of self-regulation, note taking, and test taking. *Journal of Educational Psychology, 95*(2), 335–346. https://doi.org/10.1037/0022-0663.95.2.335.

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., 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.

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., VandeKamp, K. O., Mercer, L. K., & Nordby, C. J. (2002). Investigating teacher-student interactions that foster self-regulated learning. *Educational Psychologist, 37*(1), 5–15.

Pospeschill, M. (2010). *Testtheorie, Testkonstruktion, Testevaluation [Test theory, test construction, test evaluation]*. Stuttgart: Reinhardt.

Pressley, M., Borkowski, J. G., & Schneider, W. (1987). Cognitive strategies: Good strategy users coordinate metacognition and knowledge. *Annals of Child Development, 4,* 89–129.

Ramm, G., Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., et al. (Eds.). (2006). *PISA2003: Dokumentation der Erhebungsinstrumente [Pisa 2003: Documentation of the assessment instruments]*. Waxmann: Münster.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. https://doi.org/10.1037/a0026838.

Roth, A., Ogrin, S., & Schmitz, B. (2016). Assessing self-regulated learning in higher education: A systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability, 28*(3), 225–250.

Rovers, S. F. E., Clarebout, G., Savelberg, H. H. C. M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2019). Granularity matters: Comparing different ways of measuring self-regulated learning. *Metacognition and Learning, 14*, 1–19. https://doi.org/10.1007/s11409-019-09188-6.

Schiefele, U. (1996). *Motivation und Lernen mit Texten [Motivation and learning with texts]*. Göttingen: Hogrefe.

Schiefele, U., & Schaffner, E. (2015). Motivation. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie [Educational Psychology]* (pp. 153–175). Berlin: Springer.

Schlagmüller, M., & Schneider, W. (2007). Würzburger Lesestrategie Wissenstest für die Klassen 7–12 [Würzburg reading strategy knowledge test for the grades 7–12]. In M. Hasselhorn, H. Marx & W. Schneider (Eds.) *Deutsche Schultests* [German school tests]. Göttingen: Hogrefe.

Schmidt, K., Allgaier, A., Lachner, A., Stucke, B., Rey, S., Frömmel, C., et al. (2011). Diagnostik und Förderung selbstregulierten Lernens durch Self-Monitoring-Tagebücher [Diagnostic and fostering of self-regulated learning via self-monitoring diaries]. *Zeitschrift für Hochschulentwicklung, 6*(3), 246–269.

Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik [Psychological assessments]*. Berlin: Springer.

Schmitz, B., & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning. Time-series analyses of diary data. *Contemporary Educational Psychology, 31*(1), 64–96. https://doi.org/10.1016/j.cedpsych.2005.02.002.

Schwinger, M., von der Laden, T., & Spinath, B. (2007). Strategien zur Motivationsregulation und ihre Erfassung [Strategies for motivation regulation and their assessment]. *Zeitschrift für Entwicklungspsychologie und Pädagogische, 39*(2), 57–69.

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin, 137*(3), 421–442. https://doi.org/10.1037/a0022777.

Spörer, N., & Brunstein, J. C. (2006). Erfassung selbstregulierten Lernens mit Selbstberichtsverfahren: Ein Überblick zum Stand der Forschung [Assessment of self-regulated learning via self-reports: An overview to the state of research]. *Zeitschrift für Pädagogische Psychologie, 20*(3), 147–160.

Throndsen, I. (2011). Self-regulated learning of basic arithmetic skills: A longitudinal study. *British Journal of Educational Psychology, 81*(4), 558–578. https://doi.org/10.1348/2044-8279.002008.

Trevors, G., Feyzi-Behnagh, R., Azevedo, R., & Bouchet, F. (2016). Self-regulated learning processes vary as a function of epistemic beliefs and contexts: Mixed method evidence from eye tracking and concurrent and retrospective reports. *Learning and Instruction, 42,* 31–46. https://doi.org/10.1016/j.learninstruc.2015.11.003.

Tynjälä, P., Salminen, R. T., Sutela, T., Nuutinen, A., & Pitkänen, S. (2005). Factors related to study success in engineering education. *European Journal of Engineering Education, 30*(2), 221–231. https://doi.org/10.1080/03043790500087225.

Veenman, M. V. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning, 6*(2), 205–211. https://doi.org/10.1007/s11409-011-9080-x.

Veenman, M. V., Prins, F. J., & Verheij, J. (2003). Learning styles: Self-reports versus thinking-aloud measures. *British Journal of Educational Psychology, 73*(3), 357–372.

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills. Age and task differences. *Learning and Individual Differences, 15*(2), 159–176. https://doi.org/10.1016/j.lindif.2004.12.001.

Veenman, M. V., & van Cleef, D. (2018). Measuring metacognitive skills for mathematics: students' self-reports versus on-line assessment methods. *ZDM Mathematics Education, 51,* 691–701.

Wei, M., Russell, D. W., & Zakalik, R. A. (2005). Adult attachment, social self-efficacy, self-disclosure, loneliness, and subsequent depression for freshman college students: A longitudinal study. *Journal of Counseling Psychology, 52*(4), 602–614. https://doi.org/10.1037/0022-0167.52.4.602.

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., & Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning, 4*(1), 63–85.

Wild, K.-P., & Schiefele, U. (1994). Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens [Learning strategies in study: Results concerning the factor structure and reliability of a new questionnaire]. *Zeitschrift für Differenzielle und Diagnostische Psychologie, 15,* 185–200.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*(4), 267–276.

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. 532–568). San Diego: Academic Press.

Wirth, J., & Leutner, D. (2008). Self-regulated learning as a competence: Implications of theoretical models for assessment methods. *Zeitschrift für Psychologie, 216*(2), 102–110. https://doi.org/10.1027/0044-3409.216.2.102.

Wolters, C. A. (2003). Regulation of motivation: Evaluating an underemphasized aspect of self-regulated learning. *Educational Psychologist, 38*(4), 189–205. https://doi.org/10.1207/S15326985EP3804_1.

Wolters, C. A., & Won, S. (2018). Validity and the use of self-report questionnaires to assess self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 307–322). New York: Routledge.

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. 13–41). San Diego: Academic Press.

Zimmerman, B. J. (2011). Motivational sources and outcomes of self-regulated learning and performance. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 49–64). New York: Routledge.

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modeling, 53*(1), 108–127.

Züll, C. (2014). *Offene Fragen [Open-ended questions]*. Mannheim: GESIS—Leibniz Institut für Sozialwissenschaften. https://doi.org/10.15465/sdm-sg_002.

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