Original Article

Metacognitive Judgment Skills and the Metacognitive Component of Self-Regulated Learning

A Person-Oriented, Multimethod Approach

Laura Dörrenbächer-Ulrich and Franziska Perels

Department of Educational Sciences, Faculty for Human and Economical Sciences, Saarland University, Saarbrücken, Germany

Abstract: Metacognitive judgments as part of metacognitive monitoring can be measured using different methods and show individual differences. Moreover, metacognitive judgments are highly related to the metacognitive component of self-regulated learning (SRL-MC). Nevertheless, it is unclear how metacognitive judgments are related to different ways of measuring SRL-MC. Adopting a person-oriented, multimethod approach, we measured three metacognitive judgment forms in a sample of 99 college students. Latent profile analyses resulted in four groups with differing profiles of these metacognitive judgment measures. Linking the profiles to performance, we could contextualize them within the unskilled- and unaware-effect and extend previous research on this effect. Regarding their relationship to SRL-MC, we found no differences for questionnaire values, microanalysis results, and strategy knowledge scores. These results are discussed with regard to the conceptual overlap of metacognitive judgments and SRL-MC.

Keywords: metacognitive monitoring, metacognitive judgments, self-regulated learning, latent profile analyses, unskilled and unaware effect

Metakognitive Urteilsfähigkeit und die metakognitive Komponente selbstregulierten Lernens. Ein personenorientierter, multimethodaler Ansatz

Zusammenfassung: Metakognitive Urteilsfähigkeit als eine Komponente metakognitiven Monitorings kann über verschiedene Erfassungsmethoden gemessen werden und zeigt individuelle Unterschiede. Darüberhinaus steht die metakogntive Urteilsfähigkeit in engem Zusammenhang zur metakognitiven Komponente selbstregulierten Lernens (SRL-MC). Es ist jedoch unklar, wie die metakognitive Urteilsfähigkeit mit verschiedenen Methoden zur Erfassung von SRL-MC zusammenhängt. Im Rahmen eines personenorientierten multimethodalen Ansatzes wurden drei Formen der metakognitiven Urteilsfähigkeit bei *N* = 99 Studierenden erhoben. Latente Profilanalysen ergaben vier Gruppen, die sich im Hinblick auf die Formen der metakognitiven Urteilsfähigkeit sowie Leistung unterscheiden. Diese Profile können in Bezug auf den „unskilled and unaware“-Effekt kontextualisiert werden und erweiteren den bisherigen Forschungsstand zu diesem Effekt. Im Hinblick auf SRL-MC zeigten sich keine Unterschiede zwischen den Profilen in Fragebogenwerten, mikroanalytischen Erfassungen oder Scores des Strategiewissens. Die Ergebnisse werden im Hinblick auf die konzeptuelle Überschneidung von metakognitiver Urteilsfähigkeit und SRL-MC diskutiert.

Schlüsselwörter: metakognitives Monitoring, metakognitive Urteilsfähigkeit, selbstreguliertes Lernen, latente Profilanalysen, unskilled and unaware-Effekt

According to Pintrich et al. (2000), metacognition is made up of metacognitive knowledge, metacognitive monitoring and judgments, and metacognitive control. While metacognitive knowledge refers to declarative (“what”), procedural (“how”), and conditional (“when” and “why”) strategy knowledge, metacognitive monitoring refers to the learner’s awareness concerning his/her current state of knowledge or the evaluation of a learning process in the sense of comprehension and performance (Dunlosky & Metcalfe, 2008). It is seen as a “situation-specific and context-dependent process” on the object level and provides information for the learner to regulate the learning process on the meta level (Händel & Dresel, 2022, p. 2; Nelson & Narens, 1990). Metacognitive judgments (i.e., a probabilistic judgment on the quality of performance) are seen as indicators for metacognitive monitoring (Tarricone, 2011). Metacognitive control covers the selection and usage of strategies to correct non-constructive learning pathways. Although metacognitive monitoring has been widely researched in the past few years, only a small number of studies investigated individual differences regarding this aspect of metacognition (Händel et al., 2020). Moreover, they mostly only used one type of metacognitive judgment. As a consequence, results on how different metacognitive judgment forms are related and the investigation of "metacognitive types" are sparse or nonexistent.

While the aforementioned definition on metacognitive monitoring represents an important construct of the traditional research line on metacognition, metacognition is also considered a central component of self-regulated learning (SRL; Pintrich et al., 2000). SRL is defined as "processes whereby learners personally activate and sustain cognitions, affects, and behaviours that are systematically oriented towards the attainment of personal goals" (Zimmerman, 2011, p. 1) and comprises cognitive and motivational components besides the metacognitive component (Perels et al., 2020). Adopting an SRL lens on metacognition, it can be seen as a broader competence that helps learners "to guide and direct their own learning process" (Boekaerts, 1999, p. 451) and includes planning, performing, monitoring, evaluating, and regulating one's own learning process. Therefore, the metacognitive component of SRL refers to all phases of a learning process (Zimmerman, 2000).

It is obvious that research lines on metacognition and on the metacognitive component of SRL ("SRL-MC" in the following) share a lot of commonalities and are hard to differentiate in some aspects. With regard to metacognitive monitoring, Händel and Dresel (2022) outline that it is a central factor both of SRL and metacognition models but that the constructs have rarely been investigated in combination. The authors found that SRL monitoring strategies and monitoring judgment accuracy were distinct components of an integrated metacognitive monitoring model. However, the study focused on metacognitive monitoring and did not take the whole SRL cycle of planning, performing, and evaluating (Zimmerman, 2000) into account. Concluding, the aim of the present study was twofold: Within a person-oriented and multimethod approach, the first aim was to investigate individual profiles of metacognitive judgment skills by comparing and combining several metacognitive judgment forms. To tackle the outlined conceptual enmeshment of metacognitive monitoring and SRL-MC, the second aim was to investigate how these metacognitive judgment profiles are related to different SRL-MC measures.

Metacognitive Monitoring and Judgments

In their conceptual framework on the relationship of metacognition and cognition, Nelson and Narens (1990) define an object level (basic information processing operations) and a meta level (learner's model of the task as well as cognitive operations during performance). Both levels are cyclically connected through metacognitive monitoring processes transferring information about the object level to the meta level. In addition, control processes are initiated by the meta level and regulate object-level processes to reach a specific goal. In this context, metacognitive judgment is defined as a "probabilistic judgment of one's performance before, during, or after performance [...and] can proceed or follow the completion of a test item" (Schraw, 2009, p. 34). Forms of metacognitive judgments differ with regard to the time point when judgments are made: Judgments of learning (JOL) are made during knowledge acquisition while feelings of knowing (FOK) judgments are made when retrieving learning contents (Nelson & Narens, 1990). JOL refer to how well a learning content has been learned and how well it will be remembered in the future, which is why they can be perceived as prospective monitoring. When giving prospective judgments, people use general knowledge concerning their memory functioning as well as experiences with the specific task type (Siedlecka et al., 2016). By contrast, FOK judgments mostly refer to the correctness of test item answers where learners should rate the degree of confidence they have concerning their test performance. FOK judgments are always given after performance (retrospective monitoring) and can be local (confidence for single test items) or global (confidence for overall test performance; Händel et al., 2020). By comparing confidence judgments with actual performance, researchers can compute judgment accuracy or calibration scores. Both JOL and judgment accuracy have been shown to be positively related to performance (e.g., Chua et al., 2009; Nietfeld et al., 2005). Nevertheless, learners mostly are not able to accurately judge their performance (Bol & Hacker, 2012) and tend to overestimate performance (e.g., Händel & Dresel, 2018), while overestimation seems to be associated with poorer academic achievement (Dunlosky & Rawson, 2012). Due to the multiplicity of metacognitive judgment forms caused by timing and granularity, Schraw (2009) argues for the use of multiple measures to cover different facets of metacognitive monitoring (see also Händel et al., 2020).

Individual Differences in Metacognitive Judgments

Research has shown that learners differ with regard to their metacognitive judgment skills: A common finding is that low-performing students are overconfident concerning their achievement, which has been found consistently across domains and judgment types (Bol & Hacker, 2012). This so-called unskilled and unaware effect is explained mainly by the fact that the same knowledge/abilities are necessary to answer a test item and to judge the appropriateness of a solution (Kruger & Dunning, 1999). By contrast, high-performing students tend to underestimate their achievement, but this finding is not that consistent (Erickson & Heit, 2015). Although the unskilled and unaware effect has been largely replicated, Miller and Geraci (2011) found that unskilled students also can be subjectively aware. In line with this, Urban and Urban (2021) found three types of learners: unskilled and unaware learners (low-performance, overestimation), skilled and unaware learners (high-performance, underestimation), and unskilled but aware learners (low-performance, accurate estimation). As most studies investigated the aforementioned effect using global judgments, Händel and Dresel (2018) included local judgments since the mechanisms for global and local judgments might differ. The authors found that low-performing students were actually aware of their missing knowledge with regard to incorrect items that were judged as incorrect. By contrast, high-performing students were shown to be unaware when they missed the correct answer. The authors conclude that results differ depending on the judgment form.

As a theoretical rationale, Händel et al. (2020) used the cue-utilization approach by Koriat (1997) to explain metacognitive judgments: It is assumed that global metacognitive judgments are strongly influenced by information-based cues (e.g., domain-familiarity, self-concept) that are seen as rather stable. By contrast, local judgments should be influenced by experienced-based cues concerning the concrete task. While prospective judgments only can use information-based cues, retrospective judgments can use information-based and experience-based cues and have been found to show higher accuracy. Händel et al. (2020) examined the bases for students' judgments and how they contribute to individual differences in metacognitive judgments: They found that performance and domain-specific self-concept predict metacognitive judgments. Moreover, they found that motivational and personality variables (Händel et al., 2020) influence metacognitive judgments. Concluding, it can be said that although there have been several studies on individual differences in metacognitive judgments, it is not clear how differing types of judgments influence these differences.

Metacognitive Monitoring and SRL-MC

Although metacognitive monitoring and SRL-MC share a lot of commonalities, the constructs are rarely investigated in combination (Händel & Dresel, 2022). When connecting metacognitive monitoring and SRL-MC, it can be helpful to look at SRL process models (Zimmerman, 2000): Monitoring provides learners with information on how to proceed regarding a learning goal or a specific task instruction and it enables learners to recognize needs for regulation (Tuysuzoglu & Greene, 2015). Control processes comprise the use of adequate learning strategies to pursue goal achievement and to regulate one's own behavior if necessary (Hadwin & Webster, 2013). Besides monitoring and control, Wang (2015) adds the metacognitive processes of planning and evaluation to connect both theories: While planning involves goal setting and learning strategy selection, evaluation is based on judgments of goal achievement and helps to adapt future learning processes by using experiences made in previous learning processes. It is obvious that metacognitive processes are highly relevant to all phases of SRL. In line with this, Bol and Hacker (2012) hypothesized that higher metacognitive judgment accuracy results in greater potential for the use of metacognitive SRL strategies as learners detect knowledge gaps and use metacognitive SRL strategies to regulate the goal-achieving process. Händel and Dresel (2022) state that an integrated approach is needed that investigates metacognitive monitoring judgments and monitoring strategies (as they are used in SRL research) in common (for a discussion, see also Azevedo, 2009). The authors validated a theoretical model of metacognitive monitoring composed of monitoring strategy use and metacognitive judgment accuracy. Concerning measurement of metacognitive monitoring and SRL-MC, Dinsmore et al. (2008) identified a key difference, which is task involvement. While monitoring accuracy measurement contextualizes learners' judgments of their performance within a learning task and relates it to actual performance outcomes, SRL-MC measurements oftentimes are decontextualized without referring to a specific task (e.g., questionnaires). The following section will give a short overview of common assessment methods for SRL-MC.

Assessment of SRL-MC

When assessing SRL (and SRL-MC), researchers can use different methods based on the underlying theoretical framework (Rovers et al., 2019). *Aptitude (offline)* measures assess general, relatively stable learning behavior tendencies or competences needed for SRL, whereas *event (online)* measures assess behavior that is dependent on specific learning situations and therefore changes in response to time and situations (Cleary & Callan, 2018). While offline measures often mix up strategy knowledge and strategy usage (Artelt, 2000), online measures help to capture real-time strategy usage (Schunk & Greene, 2018). Wirth and Leutner (2008) moreover differentiate between *quantitative standards* ("maximum view," the more strategies, the better the performance) and *qualita-*

tive standards ("optimum view," the better the fit between strategies and situation, the better the performance). In the following, three of the most widely used forms of assessment in SRL research are described. It will become obvious that all kinds of assessment have benefits and points of criticism. This is why several authors (e.g., Rovers et al., 2019) argue for a combination within a multimethod approach to measure SRL.

Self-report questionnaires reflect quantitative offline measures and aim to capture general learning behavior. Higher item agreement is hypothesized to predict learning outcomes. Although they can be used economically and can be standardized, they have been largely criticized (Rovers et al., 2019): Since SRL-MC is assessed as a context-independent trait, the context for answering the items is unclear. Moreover, retrospective and aggregated reports on strategy usage can be very complex, which is why generalizations or retention problems are likely. To solely asses the knowledge component of SRL-MC, researchers can use *strategy knowledge tests*, which represent qualitative offline instruments (knowledge about appropriateness of specific strategies for specific situations in the sense of conditional strategy knowledge; Wirth & Leutner, 2008). Learners receive representative learning scenarios and have to rate the effectiveness of different learning strategies for the given situation (Händel et al., 2013). Strategy knowledge tests can also be used economically and show high objectivity, while being predictive of learning performance (Maag Merki et al., 2013). *Microanalytic assessment,* which is seen as a qualitative online measure, aims at capturing specific processes and strategy usage during a learning process in a fine-grained way (Cleary & Callan, 2018). Learners have to answer questions during the phases of planning, performing, and reflection of a learning process. This method has the benefit of not being retrospectively biased (Cleary et al., 2012). Although microanalytic assessment shows low convergence to SRL questionnaires (Dörrenbächer-Ulrich et al., 2021), they seem appropriate for predicting performance and show good validity and reliability (Cleary et al., 2012).

Relationships Between Metacognitive Judgments and SRL-MC

In general, it is hypothesized that greater metacognitive judgment accuracy results in greater potential for the use of SRL strategies (Bol & Hacker, 2012). If learners can detect knowledge gaps, they should be able to use SRL strategies to regulate the goal-achieving process. To date, only few studies have investigated the relationship between metacognitive judgments and SRL-MC. Sperling et al. (2004) as well as Saraç and Karakelle (2012) report nonsignificant correlations between a metacognition questionnaire and monitoring accuracy. The latter authors explain these findings by using Koriat's theory (2000) on knowledge-based metacognitive judgments that refer to explicit and conscious inferential processes, while experience-based judgments stem from subjective feelings. They consider offline methods (e.g., questionnaires) to rely on explicit inferential processes while online methods (e.g., metacognitive judgments) use more implicit and unconscious processes. In the study by Händel and Dresel (2022), the quantity of monitoring strategy use (assessed by an offline self-report measurement) was also not related to judgment accuracy when framed in the context of general exam preparation, but both constructs were weakly related when framed to a concurrent study situation. In line with this, Veenman (2005) concludes in his review that SRL-MC offline measures do not correspond to actual and task-specific online measures.

Concerning online assessments of SRL-MC, Saraç and Karakelle (2012) found a negative correlation between metacognitive think-aloud protocols and monitoring accuracy The authors explain the negative correlation using the underconfidence-with-practice effect (monitoring accuracy decreases with increasing practice; Koriat et al., 2002), as participants in think-aloud tasks can practice until mastery, making them less accurate. Regarding metacognitive knowledge, Händel and Dresel (2022) found a low correlation with quantity of monitoring strategy use and no correlation with judgment accuracy when measured in a general context, but moderate correlations with both constructs when measured in a situation-specific way. As Griffin et al. (2013) state, metacognitive knowledge can influence the learning process without the learner engaging in online monitoring or regulation. The authors assume that a priori strategy selection (which is based on metacognitive knowledge) can even hinder online strategy regulation due to an overly strong strategy commitment.

The Present Study

Metacognitive monitoring is an important skill in the context of SRL and helps learners to cope with learning problems by using SRL strategies (Hadwin & Webster, 2013). Accordingly, learners with high metacognitive judgment accuracy show higher performance levels (Nietfeld et al., 2005). Several forms of metacognitive judgments exist, all shedding light on different aspects of monitoring. With regard to individual differences in metacognitive judgments, existing studies have focused on personality variables and achievement (Bol & Hacker, 2012; Händel et al., 2020) but to our knowledge have not investigated how differing forms of metacognitive judgments are related within one person. Moreover, higher metacognitive judgment accuracy should be related to a more ideal way of using SRL-MC strategies (Händel & Dresel, 2022). As SRL-MC can be assessed using different methods, it would be interesting to know whether the influence of metacognitive judgment accuracy on SRL-MC is dependent on the method of measurement.

Summarizing, the aim of our study was twofold: In the first step, we investigated individual differences in metacognitive judgment skills by using multiple measures for metacognitive judgment and thereby identified types of learners. In order to contextualize our findings with regard to the unskilled and unaware effect, we looked at performance differences between these groups. In the second step, we aimed to examine differences between the metacognitive judgment types with regard to SRL-MC concerning all phases of a self-regulated learning cycle (Zimmerman, 2000) measured by multiple methods. On the basis of previous results, we hypothesized there would be higher strategy use for learners with high metacognitive judgment accuracy in an SRL-MC online measure that captures actual strategy use and with regard to SRL-MC strategy knowledge. With regard to a highly generalized and retrospective questionnaire, we did not expect to find differences that are dependent on the metacognitive judgment skills.

Method

Sample and Procedure

The sample consisted of 99 college students of a southwestern German university (female = 75.8 %; age: M = 21.33, SD = 3.91), with 62 being student teachers, 36 being psychology students, and one student did not indicate study subject. We recruited student teachers and psychology students as both groups had to collect test person hours for their curriculum. Power analysis for multivariate analysis of variance (conducted to investigate differences in SRL measures between the profile groups) resulted in a sample size of N = 92 to detect small-to-midsize effects of f^2 = .06 (Cohen, 1988) with α = .05 and β = .80. For the performance measure, we conducted univariate analysis of variance. Power analysis resulted in a sample size of N = 76 to detect large effects of f = .40 (Cohen, 1988; we assumed large effects due to the existing literature on the unskilled and unaware effect) with α = .05 and β = .80. The average college entrance diploma grade (*Abitur*) was M = 1.92 (SD = .63, *range* = 1.0–3.5), with lower grades indicating higher performance. Although student teachers and psychology students significantly differed in their college entrance diploma grade, $F(1, 96) = 11.01$, $p < .01$, which is due to grade-based study admission restrictions for psychology, we found no significant differences between both groups for any of the metacognitive judgment variables, metacognitive SRL variables, or text knowledge test score. The students were in their second semester of study (M = 1.51, SD = 1.37, range = 1–7). All tests were conducted in computerized and standardized online settings. Data acquisition was completely anonymized. The participants had to sign informed consent, and were rewarded with test person hours in compensation. Students were given a link to start the survey. First, they had to give information on demographic variables. After that, they worked on a nonfictional text with the aim of preparing the information to answer a corresponding knowledge test with metacognitive judgments. In this context, microanalytic questions were implemented. After the text knowledge test, a strategy knowledge test on SRL was presented, followed by an SRL self-report questionnaire. Overall, participants worked between 60 and 90 min to complete the survey.

Instruments

Text Reading Task and Knowledge Test

Participants had to learn a text about "Sepak Takraw," an Asian ball sport, as we assumed pretest knowledge on this topic to be low. The text comprised 1,214 words and had a readability index of 44.35 (LIX; Lenhard & Lenhard, 2014–2022), which is rather low in complexity. Due to this low complexity, students only had 6 min to learn and prepare the text for a knowledge test in order to make the task challenging. The knowledge test comprised 15 multiple-choice questions on factual knowledge reported in the text (e.g., "Which material is the ball for Sepak Takraw made of?"). Each item had four answer options, one of them being correct (e.g., "leather," "plastic," "rattan," "felt"). For each correct answer, students gained 1 point. We assumed content validity to be given as the items covered the whole text and its contents. Difficulties ranged between p = .23 and p = .95 with a mean difficulty of p = .69, indicating that the test was rather easy. The mean sum score was M = 10.31 (SD = 2.51, range = 2–15), not indicating a ceiling effect. Reliability of the test (Cronbach's α) was α = .62. As a hint for criterion validity, we looked into correlations with metacognitive judgments (see below), finding a negative correlation with global prospective bias (r = –.46, p < .001) and local retrospective bias (r = –.80, p < .001), and a positive correlation with judgments of learning (r = .33, p < .001). This is in line with previous results on the relationship of achieve-

Table 1. Descriptive statistics for metacognitive judgments and SRL-MC measures

measure	*M* (*SD*)	range	theoretical Min/Max	Cronbach's alpha (items)
MJ-JOL-PQ	2.86 (0.48)	1.67 – 4.0	0/4	.81 (3)
MJ-GPB	2.64 (2.53)	-3 – 10	-15/15	–
MJ-LRB	0.04 (0.15)	-0.40 – 0.57	-1/1	–
Q-MC	2.91 (0.45)	1.74 – 3.89	0/4	.90 (19)
planning	2.62 (0.74)	1.00 – 4.00	0/4	.85 (5)
self-monitoring	3.05 (0.44)	1.80 – 4.00	0/4	.70 (5)
self-evaluation	2.98 (0.57)	1.00 – 4.00	0/4	.70 (4)
self-reaction	2.96 (0.49)	1.75 – 4.00	0/4	.68 (4)
SKT-MC	34.71 (7.60)	14 – 51	0/54	.81 (27)
planning	13.87 (3.72)	2 – 18	0/18	.81 (9)
self-monitoring	13.51 (3.55)	4 – 18	0/18	.79 (9)
self-evaluation	7.42 (4.29)	0 – 16	0/18	.84 (9)
MA-MC	7.49 (3.04)	-3.50 – 14.00	-6/20	(8)[a]
planning	-0.10 (1.22)	-1.00 – 3.00	-1/n.a.[b]	
self-monitoring	5.62 (2.07)	-1.50 – 10.50	-5/n.a.[b]	
self-reaction	1.97 (1.21)	-2.00 – 4.00	-2/4	

Note. MJ = metacognitive judgment, JOL-PQ = judgment of learning process quality, GPB = global prospective bias, LRB = local retrospective bias, Q = questionnaire, SKT = strategy knowledge test, MA = microanalysis, MC = metacognition. [a]For microanalysis, it is not reasonable to calculate internal consistencies as the scores of the coding scheme are not consistent across questions. [b]It is not possible to indicate a theoretical maximum, as participants were given one point per strategy they named (which was not limited).

ment and metacognitive judgments (e.g., Dunlosky & Rawson, 2012).

Metacognitive Judgments

Global Judgment on Learning Process Quality (JOL-PQ)

After the learning time, we asked students to subjectively indicate the quality of their learning process by asking: (1) "How well did you understand the text?"; (2) "How well could you learn the text?"; and (3) "How well will you do in the knowledge test?" All questions should be answered on a 4-point scale ranging from 1 = *totally not sure/well* to 4 = *totally sure/well.* We built up a mean score of all three questions. This newly built scale for global judgement of learning process quality had a reliability of Cronbach's α = .81, with interrelation between the three questions ranging from .50 < r < .70. For descriptive statistics see Table 1.

Global Prospective Bias (GPB)

After the question on learning process quality, we asked, "How many of the 15 questions will you answer correctly in the knowledge test?" to get a more detailed prospective performance judgment. By calculating the difference between self-judged prospective and later actual performance, we obtained a global prospective bias value. The score ranges from –15 to 15, with scores below 0 representing underconfidence and scores above 0 representing overconfidence. For descriptive statistics, see Table 1.

Local Retrospective Bias (LRB)

After each test item, students were asked to indicate how certain they feel (FOK) with the selected answer on a 4-point scale from 1 = *very uncertain* to 4 = *very certain.* A bias score for local judgments was created by converting the 4-point scale into percentage scores, as suggested by Schraw (2009). We then calculated individual bias scores by using the formula of Schraw (2009; see below). Bias scores range from –1 to 1, with scores below 0 representing underconfidence and scores above 0 representing overconfidence. For descriptive statistics, see Table 1.

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(c_i - p_i)$$

Measures for SRL-MC

Questionnaire

The metacognitive items of an established self-report questionnaire (Dörrenbächer & Perels, 2016) were used to assess SRL-MC with a quantitative offline method. The measure is based on Zimmerman's (2000) SRL model and consists of 18 items for the metacognitive component on planning (e.g., "Before learning, I write a time schedule"), self-monitoring (e.g., "During learning, I ponder whether my course of action is wise"), self-evaluation (e.g., "After learning, I check if I have reached my goals"), and self-reaction (e.g., "After learning, I think about what

I could do better next time"). All items had to be rated on a 4-point rating scale (1 = *not true at all*, 4 = *totally true*). Confirmatory factor analysis with the four subscales mentioned above showed a good overall model fit, $\chi^2(115) = 175.60$, $p < .001$, $\chi^2/df = 1.5$, comparative fit index (CFI) = .90, root mean square error of approximation (RMSEA) = .073 [.051–.094], standardized root mean square residual (SRMR) = .069. To prevent suggestive effects on the other SRL instruments, the questionnaire was applied last. Table 1 shows descriptive statistics and reliabilities.

Strategy Knowledge Test

In order to assess the SRL-MC with a qualitative offline instrument, we used a self-created, scenario-based strategy knowledge test (Dörrenbächer-Ulrich et al., 2023) that presents scenarios on all SRL phases (planning, performance, and reflection; Zimmerman, 2000). For each phase, we presented a metacognitive problem within a learning situation (e.g., "Luisa prepares for a test that is due in 3 weeks. She is interests in the learning content and has planned to work on two chapters today. Unfortunately, she has to finish other important tasks for college today and notices that her time is running out"). Participants should rate six strategies (three useful, e.g., "She should identify subject areas that she has not understood yet and focus on them," and three less useful, e.g., "She should keep up her plan for test preparation and delay the other tasks") regarding their usefulness for the specific problem on a 4-point rating scale (1 = *not useful at all*, 4 = *very useful*). It was specified that participants should not indicate how they would solve the problem but which strategy they rate the most useful independent of their own behavior. The presented strategies were based on theoretical considerations and were rated by SRL experts from the field regarding their usefulness. To prevent obviousness and to assess conditional strategy knowledge (knowledge about the fit of situations and strategies), less useful strategies were not totally useless but represented less useful strategies concerning the specific learning scenario. Moreover, we presented situations with a female or male character fitting the gender of the participant. A score for each participant was calculated by pairwise comparisons of ratings for useful and less useful strategies resulting in nine comparisons for each phase (see Händel et al., 2013). If the less useful strategy was rated as being equally useful as or more useful than the useful strategy, 0 points were given. If participants rated the useful strategy 1 point higher than the less useful strategy, 1 point was given and if participants rated the useful strategy 2 or 3 points higher than the less useful strategy, 2 points were given (see Table 1 for descriptive statistics).

Microanalytic Assessment

Microanalytic assessment was tied to the text learning task. In the beginning, participants had 90 s to get an overview of the text. Afterward, they answered a microanalytic question on the planning phase (e.g., "Do you have a plan for working on the task?). After 3 min in which participants could learn the text, they answered five microanalytic questions on the performance phase referring to monitoring (e.g., "Do you have to change something in your behavior to successfully solve the task?"). After an additional 3 min to learn the text, students answered JOL-PQ questions as well as the knowledge test with FOK judgments. After that, participants answered two microanalytic questions on the reflection phase concerning adaptive self-reaction (e.g., "What would you do when working on this task again?"). Answers were coded by two independent trained raters following a theoretically based coding scheme used in a previous study (Dörrenbächer-Ulrich et al., 2021). For interrater reliability, we calculated intraclass correlations (ICC, 2, 1) with two-way random single measure for absolute agreement indicating high interrater agreement (range = .76–.98, which can be interpreted as good to excellent; Koo & Li, 2016). For analyses, we built a sum score for metacognitive microanalytic questions (see Table 1 for descriptive statistics).

Results

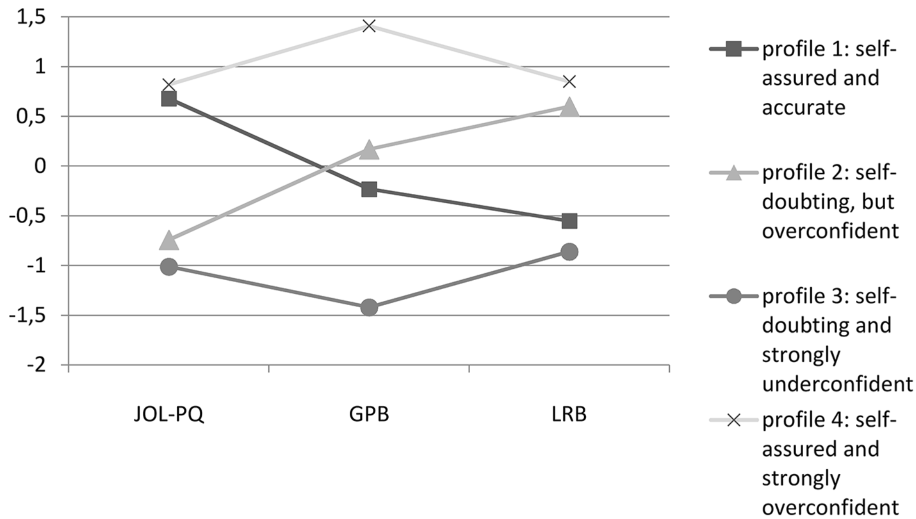
Profile Analyses of Metacognitive Judgments

For the different forms of metacognitive judgments, we found the following correlations: JOL-PQ – GPB, $r = .40$, $p < .001$; JOL-PQ – LRB, $r = .03$, $p > .05$, GPB – LRB, $r = .64$, $p < .001$. Profiles of metacognitive judgment skills were then investigated by conducting latent profile analyses (LPA; Vermunt & Magidson, 2002). LPA follows a person-oriented approach and extracts latent classes by grouping participants with similar metacognitive judgment skills into homogenous profiles. Simultaneously, LPA aims to maximize the dissimilarity between identified profiles. We conducted an exploratory LPA in MPlus7 (Muthén & Muthén, 1998–2012) by investigating models from 1–6 classes. Using a robust maximum-likelihood estimation approach (MLR), several model fit criteria helped to decide on the best-fitting latent profile model: AIC (Akaike information criterion), BIC (Bayesian information criterion), entropy (E, measure of classification certainty), BLR (bootstrapped likelihood ratio test), and the Lo–Mendell–Rubin test (LMRT; Marsh et al., 2009). We decided for the four-cluster solution, as it had the lowest

Table 2. Latent profile analysis results for the three metacognitive judgment measures

Classes	AIC	BIC	Entropy	LMRT	BLR
2	482.56	508.51	.66	.06	.00
3	470.60	506.94	.79	.18	.00
4	**458.66**	**505.37**	**.79**	**.25**	**.00**
5	450.89	507.98	.86	.24	.03
6	446.34	513.82	.88	.19	.13

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; LMRT = Lo–Mendell–Rubin test; BLR = bootstrapped likelihood ratio test. Based on the model fit indices, we decided for the 4-cluster solution, indicated in bold.



Note. JOL-PQ = judgment of learning process quality, GPB = global prospective bias, LRB = local retrospective bias.

Figure 1. Profile plots of metacognitive judgment groups.

BIC and a high entropy (see Table 2). Although the LMRT did not show a significant result, we decided on four groups as they have a good interpretability. In addition, both the five- and the six-cluster solution had classes with less than five persons (less than 5% of the sample), making them hard to interpret. Table 3 shows descriptive statistics with regard to the three metacognitive judgments. Figure 1 shows a profile plot with z-scores for the metacognitive judgment variables of the four profile groups. Using z-scores simplifies the interpretation of results and helps to gain insights on relative differences between the profiles (z-scores above .5 and below –.5 are seen as extreme; see Liu et al., 2014).

As JOL-PQ asked participants how they subjectively judge their learning process quality on a global level, we refer to the terms of "self-assured" and "self-doubting" for categorization. The first group, which we named "self-assured and accurate," had high JOL-PQ values and showed slight overconfidence prospectively and slight underconfidence retrospectively with small bias scores indicating rather accurate judgments. The second group showed low JOL-PQ values with moderate overconfidence prospectively and high overconfidence retrospectively, which is why we named this group "self-doubting, but overconfident." The third group was named "self-doubting and underconfident" and showed very low JOL-PQ values, slight underconfidence prospectively and moderate underconfidence retrospectively. The fourth group showed high JOL-PQ values and strong overconfidence prospectively and retrospectively and was named "self-assured and strongly overconfident." Univariate analyses of variance (ANOVAs) with metacognitive judgment scores as dependent variables and profile groups as independent variable revealed significant differences between the profiles. In addition, Scheffé post hoc tests showed that almost all groups differed from each other, with four exceptions listed in Table 3.

Differences Between Profiles With Regard to Performance

In the next step, we analyzed how academic performance differs between the profiles. We conducted an ANOVA with profile groups as independent variable and text knowledge test performance as dependent variable. We

Table 3. Means and standard deviations for indicator variables of the profiles

Profile	*n*	*M* (*SD*) (*Z*) JOL-PQ	*M* (*SD*) (*Z*) GPB	*M* (*SD*) (*Z*) LRB
1: Self-assured and accurate	33	3.18[a] (0.26) (0.68)	2.06 (1.17) (–0.23)	–.04[c] (0.09) (–0.55)
2: Self-doubting, but overconfident	28	2.50[b] (0.29) (–0.74)	3.07 (1.12) (0.17)	.13[d] (0.10) (0.60)
3: Self-doubting and strongly underconfident	18	2.37[b] (0.32) (–1.01)	–0.94 (1.06) (–1.42)	–.09[c] (0.11) (-0.86)
4: Self-assured and strongly overconfident	20	3.25[a] (0.32) (0.82)	6.20 (1.24) (1.41)	.17[d] (0.13) (0.85)

Note. [a,b,c,d] Same letters within one column indicate that this pairwise comparison was not significant ($p > .05$).
JOL-PQ = judgment of learning process quality; GPB = global prospective bias; LRB = local retrospective bias.

Table 4. Means and standard deviations for SRL metacognition scales and performance depending on profile group and results of the (M)ANOVA

	M (*SD*)				*F* (*df*), *p*, partial η²
	Self-assured and accurate	Self-doubting, but overconfident	Self-doubting and strongly underconfident	Self-assured and strongly overconfident	
MA_MC					$F(9, 226.49) = 1.09$, $p = .37$, partial $\eta^2 = .03$, Wilk's $\Lambda = .90$
Planning	-0.03 (1.29)	0.00 (1.23)	-0.08 (1.20)	-0.38 (1.13)	
Self-monitoring	6.06 (1.83)	5.61 (1.65)	5.75 (2.08)	4.83 (2.78)	
Self-reaction	1.91 (1.35)	2.32 (1.14)	1.56 (0.98)	1.95 (1.20)	
Q_MC					$F(12, 241.06) = 0.91$, $p = .54$, partial $\eta^2 = .04$, Wilk's $\Lambda = .89$
Planning	2.69 (0.73)	2.67 (0.62)	2.42 (0.79)	2.64 (0.86)	
Self-monitoring	3.09 (0.49)	3.04 (0.43)	2.96 (0.35)	3.08 (0.45)	
Self-evaluation	3.08 (0.49)	2.89 (0.50)	2.92 (0.72)	3.01 (0.66)	
Self-reaction	3.09 (0.47)	2.94 (0.40)	2.89 (0.52)	2.83 (0.58)	
SKT_MC					$F(9, 224.05) = 0.87$, $p = .55$, partial $\eta^2 = .03$, Wilk's $\Lambda = .92$
Planning	14.34 (3.60)	14.82 (3.36)	12.72 (2.70)	12.95 (4.87)	
Self-monitoring	13.00 (3.86)	13.32 (3.31)	14.17 (3.60)	14.20 (3.17)	
Self-reaction	7.38 (3.98)	7.32 (3.99)	7.94 (4.67)	7.15 (5.04)	
Knowledge Test score	12.09	8.50	11.39	8.95	$F(3, 95) = 22.19$, $p < .01$, $\eta p^2 = .41$

Note. MA = microanalysis; Q = questionnaire; SKT = strategy knowledge test; MC = metacognition.

found an overall difference for knowledge test score (Table 4). Scheffé post hoc tests indicated that the comparisons between Groups 1 and 2, Groups 1 and 4, and Groups 2 and 3 were significant. Therefore, the self-assured and accurate and the self-doubting and strongly underconfident groups scored best having a comparable score, while the self-doubting, but overconfident and the self-assured and strongly overconfident groups scored worst, also having a comparable score.

Differences Between Profiles With Regard to SRL-MC

Following that, we investigated how the metacognitive judgment profiles differ in terms of SRL-MC. Therefore, we analyzed differences in microanalysis scores, strategy knowledge scores, and questionnaire values by conducting MANOVAs with the metacognitive subscales as dependent variables (see Table 4). We found no significant differences between the profiles for any of the SRL-MC measures.

Discussion

The present study aimed to investigate whether there are different types of learners regarding the combination of multiple metacognitive judgment forms and to examine whether these types show differences in performance and different measures of SRL-MC. We found four groups that

Table 5. Combinations of levels of performance, subjective, and objective awareness

Group	Performance	Subjective awareness	Objective awareness
1: Self-assured and accurate	Skilled	Aware	Aware
2: Self-doubting, but overconfident	Unskilled	Aware	Unaware
3: Self-doubting and strongly underconfident	Skilled	Unaware	Unaware
4: Self-assured and strongly overconfident	Unskilled	Unaware	Unaware

differed regarding JOL-PQ as well as prospective and retrospective bias (FOK), and that showed significant performance differences. Concerning three different SRL-MC measures (two offline and one online instrument), we found no group differences.

Profiles of Metacognitive Judgment Skills

As several authors recommend the use of multiple metacognitive judgments (e.g., Händel et al., 2020), we used a JOL scale on learning process quality (JOL-PQ), a global prospective bias and a local retrospective bias to investigate metacognitive judgment types. Using LPA, we found four groups that differed regarding their JOL-PQ values as well as both prospective and retrospective bias scores: the self-assured and accurate group, the self-doubting, but overconfident group, the self-doubting and underconfident group as well as the self-assured and strongly overconfident group. We found two groups with high JOL-PQ values, that is, learners were assertive of having learned the text well and of doing well in the knowledge test. With regard to their bias values, one group was rather accurate while the other group was strongly overconfident. Moreover, we found two groups with relatively low JOL-PQ values who also differed in terms of their bias values, as one group was rather underconfident, while the other was overconfident. In order to contextualize our findings concerning the unskilled and unaware effect, we looked at performance differences between the groups. Independent of JOL-PQ level, the accurate and underconfident groups showed better results than both overconfident groups. This is in line with previous results that low-performing students tend to overestimate their performance (Händel & Dresel, 2018), while high-performing students tend to underestimate their performance. If we use the JOL-PQ measures as a form of subjective awareness and the bias scores as objective awareness, we could expand the categorization that has been used to date (see Table 5 for categorization). While Urban and Urban (2021) found an unskilled and unaware group, a skilled and unaware group, and an unskilled but aware group, we found a fourth group that is skilled and aware of it. Moreover, we uncovered a difference in subjective and objective awareness using multiple metacognitive judgment forms. This speaks in favor of a multimethod approach to measure metacognitive judgments, since subjective and objective judgments not always go hand in hand.

Differences Between Metacognitive Judgments Profiles in SRL-MC Measures

In a further step, we looked into differences between metacognitive judgment profiles with regard to different SRL-MC measures. Overall, we found no group differences concerning any of the SRL-MC measures. This was what we hypothesized for the SRL-MC questionnaire and is in line with previous results on the relationship of offline methods and monitoring accuracy (Händel & Dresel, 2022; Saraç & Karakelle, 2012; Sperling et al., 2004). As offline methods rely rather on knowledge-based metacognitive judgments, which refer to explicit and conscious inferential processes, and online methods use more implicit and unconscious processes, that is, experience-based judgements, as a base, a weak relationship seems to be justified theoretically (Saraç & Karakelle, 2012). In terms of strategy knowledge, the nonexistent differences can be explained by the fact that the strategy knowledge test was very global (referring to the general usefulness of SRL-MC strategies with regard to a given problematic situation). Therefore, the same explanation as for offline questionnaire could be used here. In line with this, Händel and Dresel (2022) also found no correlation between metacognitive judgment accuracy and strategy knowledge when measured in a global way.

Accordingly, previous research on SRL strategy knowledge in general has shown that is not closely related to strategy usage (Dörrenbächer-Ulrich et al., 2021).

An explicit unexpected finding was the missing difference between metacognitive judgment profiles with regard to actual SRL-MC strategy usage measured through microanalysis. In general, it is hypothesized that greater metacognitive judgment accuracy results in greater potential for the use of SRL strategies (Bol & Hacker, 2012) as learners can detect knowledge gaps and therefore use SRL strategies to regulate the goal-achieving process.

Händel and Dresel (2022) give a possible explanation for this result and hypothesize that the relationship between judgment accuracy and the actual use of monitoring strategies depends on performance level. If a learner is performing high and is accurate in his/her metacognitive monitoring and judgment of this high performance, he/she does not need to use a lot of metacognitive strategies (as he/she is already "on the right way"). If a learner is performing low and is accurately monitoring this, he/she should intensify the use of metacognitive learning strategies. Thus, in our sample, it could be the case that the accurate, high-performing group did not need to use more strategies. The overconfident groups did not see any need to use metacognitive strategies as they overestimated their performance and therefore did not realize that strategy usage was needed. For the underconfident group (which was actually performing high) it could be hypothesized that they saw the need to use learning strategies as they did not estimate their performance to be high. Nevertheless, since this group showed "self-doubting" JOL-PQ, it could be possible that they also had a low self-efficacy for the use of SRL strategies and therefore did not use any.

Limitations and Implications

The present study has several points of criticism that could be tackled in future studies. First, our sample size was relatively small as we used a wide range of instruments making test sessions rather long and laborious for participants. Apart from this, since open answers in microanalysis must be coded by two raters, data preparation is very time consuming and arduous. The small sample size resulted in small profile groups and missing power for group comparisons. Larger sample sizes would have enabled us to investigate subgroup differences on univariate levels. Second, our sample was rather restrictive as participants were in their first semesters of study and therefore not very experienced with university learning. It would be interesting if the same group can be found in older students to see how they differ regarding SRL. Third, although we used several metacognitive judgment measures, we mixed up global and local judgments with the time point of judgment (Händel et al., 2020). To complete the picture, it would be helpful to also assess local prospective judgments and global retrospective judgments. Lastly, the reliability of the knowledge test was not optimal and the text was rather easy to read. It could be the case that the text learning task did not provoke a real need to use SRL-MC strategies due to the missing complexity.

Future research could investigate the profiles found here in more depth, for example, if they are found for other domains or in other age groups. Moreover, it could be analyzed whether learners change the profile they belong to with growing experience by conducting latent transition analyses. With regard to SRL-MC measures, it should be investigated further why they did not show differences depending on metacognitive judgment skills (especially microanalysis). To complete the picture, future research could examine whether there are SRL-MC measures that are more or less valid (e.g., with regard to their predictive value for performance) for different metacognitive judgment profiles of students. With regard to implications for educational practitioners, it seems important to bear in mind that learners differ with regard to their metacognitive judgment accuracy and that these differences are related to performance. With regard to the overconfident groups, metacognitive trainings could be helpful to make them more accurate in their judgments. Regarding the self-doubting groups, trainings concerning their self-efficacy for SRL could be useful in strengthening their belief that they can effectively apply SRL strategies to improve their learning.

References

Artelt, C. (2000). Wie prädiktiv sind retrospektive Selbstberichte über den Gebrauch von Lernstrategien für strategisches Lernen? [How predictive are retrospective self-reports on the use of learning strategies for strategic learning?]. *Zeitschrift für Pädagogische Psychologie*, *14*(2/3), 72–84. https://doi.org/10.1024//1010-0652.14.23.72

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. https://doi.org/10.1007/s11409-009-9035-7

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

Bol, L., & Hacker, D. J. (2012). Calibration research: Where do we go from here? *Frontiers in Psychology*, *3*, 229. https://doi.org/10.3389/fpsyg.2012.00229

Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009). Neural correlates of metamemory: A comparison of feeling-of-knowing and retrospective confidence judgments. *Journal of Cognitive Neuroscience, 21*(9), 1751–1765. https://doi.org/10.1162/jocn.2009.21123

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). Routledge.

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, 2012,* 1–19. https://doi.org/10.1155/2012/428639

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*(4), 391–409. https://doi.org/10.1007/s10648-008-9083-6

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. http://dx.doi.org/10.1016/j.ijer.2016.05.010

Dörrenbächer-Ulrich, L., Sparfeldt, J., & Perels, F. (2023). *Knowing how to learn: Development and validation of a strategy knowledge test on self-regulated learning for college students.* Manuscript submitted for publication.

Dörrenbächer-Ulrich, L., Weißenfels, M., Russer, L., & Perels, F. (2021). Multimethod assessment of self-regulated learning in college students: Different methods for different components? *Instructional Science, 49*(1), 137–163. https://doi.org/10.1007/s11251-020-09533-2

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition.* Sage.

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

Erickson, S., & Heit, E. (2015). Metacognition and confidence: Comparing math to other academic subjects. *Frontiers in Psychology*, *6*, 742. https://doi.org/10.3389/fpsyg.2015.00742

Griffin, T. D., Wiley, J., & Salas, C. R. (2013). Supporting effective self-regulated learning: The critical role of monitoring. In R. Azevedo, & V. Aleven. (Eds.), *International handbook of metacognition and learning technologies* (pp. 19–34). Springer.

Hadwin, A. F., & Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning and Instruction, 24*, 37–47. https://doi.org/10.1016/j.learninstruc.2012.10.001

Händel, M., Artelt, C., & Weinert, S. (2013). Assessing metacognitive knowledge: Development and evaluation of a test instrument. *Journal for Educational Research Online, 5*(2), 162–188.

Händel, M., & Dresel, M. (2018). Confidence in performance judgment accuracy: The unskilled and unaware effect revisited. *Metacognition and Learning*, *13*(3), 265–285. https://doi.org/10.1007/s11409-018-9185-6

Händel, M., & Dresel, M. (2022). Structure, relationship, and determinants of monitoring strategies and judgment accuracy. An integrated model and evidence from two studies. *Learning and Individual Differences*, *100*, 102229. https://doi.org/10.1016/j.lindif.2022.102229

Händel, M., de Bruin, A. B., & Dresel, M. (2020). Individual differences in local and global metacognitive judgments. *Metacognition and Learning*, *15*(1), 51–75. https://doi.org/10.1007/s11409-020-09220-0

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–63. https://doi.org/10.1016/j.jcm.2016.02.012

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2000). The feeling of knowing: Some meta-theoretical implications for consciousness and control. *Consciousness and Cognition, 9*(2), 149–171 https://doi.org/10.1006/ccog.2000.0433

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131*(2), 147–162. https://doi.org/10.1037/0096-3445.131.2.147

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lenhard, W., & Lenhard, A. (2014–2022). *Berechnung des Lesbarkeitsindex LIX nach Björnson* [Calculation of the readability index LIX according to Björnson]. Psychometrica. https://doi.org/10.13140/RG.2.1.1512.3447

Liu, W. C., Wang, C. K. J., Kee, Y. H., Koh, C., Lim, B. S. C., & Chua, L. (2014). College students' motivation and learning strategies profiles and academic achievement: A self-determination theory approach. *Educational Psychology, 34*(3), 338–353. https://doi.org/10.1080/01443410.2013.785067

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. https://doi.org/10.1891/1945-8959.12.3.391

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. (2009). Classical latent profile analysis of academic self-concept dimensions. Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling, 16*(2), 191–225. https://doi.org/10.1080/10705510902751010

Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 502–506. https://doi.org/10.1037/a0021802

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide: Statistical analysis with latent variables* (7th ed.). Muthén & Muthén.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26*, 125–141. https://doi.org/10.1016/S0079-7421(08)60053-5.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education, 74*(1), 7–28.

Perels, F., Dörrenbächer-Ulrich, L., Landmann, M., Otto, B., Schnick-Vollmer, K., & Schmitz, B. (2020). Selbstregulation und selbstreguliertes Lernen. [Self-regulation and self-regulated learning]. In E. Wild, & J. Möller. (Eds.), *Pädagogische Psychologie.* (pp. 45–66). Springer. https://doi.org/10.1007/978-3-642-41291-2_3

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw, & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Buros Institute of Mental Measurement. https://digitalcommons.unl.edu/burosmetacognition/3

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), 1–19. https://doi.org/10.1007/s1140 9-019-09188 -6.

Saraç, S., & Karakelle, S. (2012). On-line and off-line assessment of metacognition. *International Electronic Journal of Elementary Education, 4*(2), 301–315.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Schunk, D. H., & Greene, J. A. (2018). *Handbook of self-regulation of learning and performance* (2nd ed.). Routledge.

Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology, 7*, 218. https://doi.org/10.3389/fpsyg.2016.00218

Sperling, R. A., Howard, B. H., Staley, R., & DuBois, N. F. (2004). Metacognition and self- regulated learning constructs. *Educational Research and Evaluation, 10*(2), 117 – 139.

Tarricone, P. (2011). *The taxonomy of metacognition.* Psychology Press.

Tuysuzoglu, B. B., & Greene, J. A. (2015). An investigation of the role of contingent metacognitive behavior in self-regulated learning. *Metacognition and Learning*, *10*(1), 77 – 98. https://doi.org/10.1007/s11409-014-9126-y

Urban, M., & Urban, K. (2021). Unskilled but aware of it? Cluster analysis of creative metacognition from preschool age to early adulthood. *Journal of Creative Behaviour, 55*(4), 937 – 945.

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

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. *Applied Latent Class Analysis, 11*, 89 – 106. https://doi.org/10.1017/cbo9780511499531.004

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*(4), 555 – 579. https://doi.org/10.1007/s11165-014-9436-8

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

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). Academic press. https://doi.org/10.1016/B978-012109890-2/50031-7

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). Routledge. https://doi.org/10.4324/9780203839010.CH4

History
Published online May 17, 2023

Funding
This project is funded by the German Research Foundation (grant number: DO 2173/2 – 1). The authors are responsible for the content of this publication.
Open access publication enabled by Saarland University and State Library.

ORCID
Laura Dörrenbächer-Ulrich
https://orcid.org/0000-0002-8361-9030

Dr. Laura Dörrenbächer-Ulrich
Fachrichtung Bildungswissenschaften
Saarland University
Campus A4.2
66123 Saarbrücken
Germany
laura.doerrenbaecher@uni-saarland.de