Original Article

Metacognitive Monitoring via Strategies and Judgments

Different Phases, Different Objects

Marion Händel[1], Nick Naujoks-Schober[2], and Markus Dresel[3]

[1]Faculty of Media, Ansbach University of Applied Sciences, Ansbach, Germany
[2]Institute of Psychology, Faculty of Humanities, Social Sciences, and Theology, Friedrich-Alexander University Erlangen-Nuremberg, Germany
[3]Department of Psychology, Faculty of Philosophy and Social Sciences, University of Augsburg, Germany

Abstract: Metacognitive monitoring is conceptualized as a situation-specific and context-dependent process that helps learners to regulate their learning. The current study builds on the idea that metacognitive monitoring can fulfil monitoring functions in different phases (when to monitor: during learning or during testing), and that it refers to several objects (what to monitor: processing or retrieval). The cross-sectional study with 184 higher-education students used a situation-specific approach and referred to students' monitoring via monitoring strategies and monitoring judgments during test preparation and test processing. Confirmatory factor analyses indicated that monitoring via strategies and judgments can be directed at different objects. In addition, monitoring different objects was more strongly correlated within the same phase than across different phases. The study results emphasize the need for an object-specific and comprehensive consideration of metacognitive monitoring via monitoring strategies and monitoring judgments.

Keywords: metacognitive monitoring, monitoring strategies, judgment bias, phases, objects of monitoring

Metakognitive Überwachung durch Strategien und Urteile. Unterschiedliche Phasen, unterschiedliche Objekte

Zusammenfassung: Metakognitive Überwachung wird als situations- und kontextspezifischer Prozess konzeptualisiert, der Lernende dabei unterstützt, ihr eigenes Lernen zu regulieren. Die vorliegende Studie basiert auf der Annahme, dass metakognitive Überwachung in verschiedenen Phasen relevant ist (Zeitpunkt der Überwachung: während der Vorbereitung oder während des Testens) und sich auf verschiedene Objekte bezieht (Gegenstand der Überwachung: Prozess oder Abruf). Die Querschnittsstudie mit 184 Studierenden untersuchte die Überwachung von Studierenden durch Überwachungsstrategien und metakognitive Urteile während der Testvorbereitung und Testbearbeitung mittels eines situationsspezifischen Ansatzes. Konfirmatorische Faktorenanalysen ergaben, dass die Überwachung via Strategien und Urteilen auf unterschiedliche Gegenstände gerichtet sein kann. Darüber hinaus war die Überwachung verschiedener Objekte innerhalb derselben Phase stärker korreliert als zwischen verschiedenen Phasen. Die Studienergebnisse unterstreichen die Notwendigkeit einer objektspezifischen und umfassenden Betrachtung metakognitiver Überwachung durch Überwachungsstrategien und metakognitive Urteile.

Schlüsselwörter: Metakognitive Überwachung, Überwachungsstrategien, Bias, Phasen, Überwachungsgegenstände

Metacognitive monitoring is a central process for self-regulated learning (Winne & Hadwin, 1998). Throughout the learning process, students can monitor a broad set of objects: They can monitor their goals and goal attainment, their strategy use and its effectiveness, their learning progress, their comprehension and learning results, as well as their motivation or emotion (Greene & Azevedo, 2007; Miele & Scholer, 2017; Pintrich, 2004; Sobocinski et al., 2020). When narrowing the scope to monitoring cognitive aspects of learning such as learning success (Winne & Hadwin, 1998), there is still a variety of different monitoring objects (Nelson & Narens, 1990), as the following case shows. For example, students might

have set a specific learning goal (e.g., to read and understand the book chapter about technical building equipment for an upcoming performance test). The students can then observe whether their learning is efficient (e.g., Are they distracted? Does the reading take more time than expected?). They can monitor whether they reach their goal of understanding, in other words, whether their learning is successful or not (e.g., whether they understand a specific concept within in the text). In addition, in the upcoming performance test, they can monitor their answer behavior (e.g., check whether they may have accidentally omitted an answer). Finally, they can monitor how well they could retrieve the learned

content in the performance test, that is, how well they performed.

Beyond different objects of monitoring, this example illustrates that metacognitive monitoring relates to different phases in the cycle of self-regulated learning; the students can monitor their procedure and their comprehension during the acquisition of learning content (preparation phase) as well as during the testing phase (cf., Nelson & Narens, 1990). In each phase, metacognitive monitoring has specific regulation functions (Malmberg et al., 2022). Metacognitive monitoring in the preparation phase can lead to a regulation of learning behaviors (e.g., invest more time, use more useful strategies). In the testing phase, metacognitive monitoring can help to identify careless mistakes or lead to specific testing strategies such as opting-out (Law et al., 2022) or answer changing (Couchman et al., 2016; Papanastasiou & Stylianou-Georgiou, 2022; Stylianou-Georgiou & Papanastasiou, 2017). Consequently, metacognitive monitoring can result in higher performance, as indicated by previous studies (Credé & Phillips, 2011; Händel, Harder, et al., 2020).

This shows the importance of metacognitive monitoring for successful learning and performance. Still, with the exception of only one empirical study (Lan, 2005), theoretical models as well as empirical research did not explicitly distinguish different objects of monitoring. On the contrary, empirical research often neglects this differentiation and treats metacognitive strategies – such as planning, monitoring, and regulation – as a whole (Broadbent et al., 2022; Craig et al., 2020; Griffin et al., 2013). Furthermore, while different phases of monitoring are defined in theoretical models (Nelson & Narens, 1990; Winne & Hadwin, 1998), empirical studies – at least in the context of monitoring strategies – largely ignore that. Regarding monitoring judgments, however, a large body of research is concerned with judgments in different phases (e.g., predictions, concurrent judgments, or post-dictions; Hacker et al., 2008). Therefore, the aim of the current study was to differentiate different phases and objects of monitoring. A further aim was to investigate how metacognitive monitoring is interrelated in different phases and regarding different objects.

Metacognitive Monitoring

Metacognitive monitoring is a situation-specific and context-dependent process (Boekaerts, 1999; Dresel et al., 2015; Händel & Dresel, 2022; Koriat, 2019; McCardle & Hadwin, 2015; Schuster et al., 2023; Wirth & Leutner, 2008). The current work focuses on metacognitive monitoring via monitoring strategies and monitoring judg-

ments. Monitoring strategies and monitoring judgments are both aspects of metacognitive monitoring, as, for example, illustrated in the model of self-regulated learning by Winne and Hadwin (1998), where monitoring takes a central role. Monitoring strategies are – in addition to planning and regulation strategies – metacognitive strategies (Alexander et al., 1998; Pintrich, 1999). Monitoring strategies are cognitive or behavioral approaches with the aim of evaluating learning behavior and comprehension (e.g., evaluating the effectiveness of undertaken strategies, or self-explaining; Flavell, 1979). Monitoring judgments are evaluations of personal performance based on academic tasks (Koriat, 2019; Schraw, 2009b; Winne & Muis, 2011). For example, a metacognitive judgment is generated if a student thinks that he or she answered a task correctly during test processing. From a metacognitive perspective, monitoring judgments should be as accurate as possible because accurate judgments help learners to regulate their learning according to their strengths and weaknesses. Different indices provide information about judgment accuracy. For example, bias as the difference between actual performance and the performance judgment indicates whether students over- or underestimate their performance (Schraw, 2009a). If students can accurately estimate whether they have correctly solved a task, this can help them to further plan and regulate the learning process.

So far, the research traditions focusing on monitoring strategies on the one hand and monitoring judgments on the other hand were quite detached from each other. However, Händel and Dresel (2022) recently proposed an integrated model that combines two strands of research: the strategy-oriented perspective and the judgment-oriented perspective of metacognitive monitoring. The model suggests that monitoring strategies as well as judgment accuracy contribute to metacognitive monitoring and that the use of monitoring strategies enhances judgment accuracy. Two empirical studies confirmed the proposed factorial structure and revealed that the use of monitoring strategies and judgment accuracy display two – however, only slightly correlated – factors. This pattern of results holds even if monitoring strategies and monitoring judgments relate to the same situation, namely, an ungraded performance test, as in Study 2 by Händel and Dresel. These low correlations were also found in two other current studies (Dörrenbächer-Ulrich et al., 2021; Golke et al., 2022). However, to date, research did not consider that monitoring strategies and monitoring judgments can fulfil monitoring functions in different phases (when to monitor) and refer to several objects (what to monitor).

Table 1. Monitoring four different objects in two phases

| | Preparation phase | | | Testing phase | |
	Processing	Retrieval		Processing	Retrieval
Definition	Monitoring effectiveness of the learning approach	Monitoring comprehension		Monitoring test processing	Monitoring correctness of given answers
Sample monitoring behavior	Check whether the learning procedure used is effective	Check whether the learning content was understood		Check whether all items have been answered	Check whether an item answer is correct

Phases of Monitoring – When to Monitor

Metacognitive monitoring can occur during different phases in the process of self-regulated learning. In Winne and Hadwin's model (1998), monitoring relates to all phases. For example, during studying, a student can monitor his or her learning procedure and his or her comprehension. After studying, a student can monitor his or her performance (i.e., undertake a cognitive evaluation). This is in line with the model by Nelson and Narens (1990) that distinguishes monitoring judgments with regard to the phases of acquisition, retention, and retrieval of information. According to each phase, students can use different judgments to monitor their degree of knowledge. In detail, Hacker and Bol (2019) relate different types of judgments to different phases of learning. So-called predictions resemble expectation of future performance regarding a comprehension test (Hacker et al., 2000; Maki & McGuire, 2002; Miller & Geraci, 2011). During learning, concurrent judgments, usually task-specific judgments, are generated (Schraw, 2009b). During or after testing, so-called postdictions are implemented to monitor test or item performance (Händel, de Bruin, et al., 2020).

In the following, we structure metacognitive monitoring regarding two phases: First, the preparation phase, where students acquire and retain new information. Second, the testing phase, in which students generate learning products and need to retrieve knowledge (see Table 1).

Objects of Monitoring – What to Monitor

Within and across phases, students can monitor each different objects (Nelson & Narens, 1990). Students' monitoring can relate to either their processing or their retrieval and they can implement this in preparation and testing phases. Monitoring processing is usually conducted via monitoring strategies. Students can implement monitoring strategies regarding their learning or test answering approach (Stebner et al., 2022). Hence, during both the preparation and the testing phase students can

monitor the effectiveness of their studying or their testing approach via monitoring strategies. In addition, students can monitor their comprehension during the preparation phase or monitor their performance during the testing phase, that is, they monitor the outcomes of the processes. Metacognitive monitoring regarding retrieval means to monitor one's comprehension (e.g., de Bruin et al., 2016) or one's test performance displayed by given answers (Händel, Harder, et al., 2020). This is usually conducted via the generation of monitoring judgments. When monitoring comprehension in the preparation phase, students self-select the content to which they refer their comprehension monitoring. By contrast, when monitoring performance in a testing situation, students answer external questions to which they provide metacognitive judgments.

First, in the preparation phase, students use metacognitive monitoring to prove whether their learning behavior, that is, the implemented strategies and procedures, is useful to reach their goal. A strategy to monitor one's learning progress can be to pause to reflect whether the currently used strategy is effective. For example, students can ask themselves if they implement adequate and helpful learning strategies, or if they constantly engage in mindless repetition. Second, students can generate metacognitive judgments regarding their comprehension in the preparation phase after engaging in quizzing. Third, during test processing, students monitor whether their answering procedure is effective, whether their time allocation for the individual tasks is reasonable, or whether they have left out any questions (Papanastasiou & Stylianou-Georgiou, 2022; Stylianou-Georgiou & Papanastasiou, 2017). Fourth, to monitor test performance, several types of monitoring judgments can be used. The monitoring object can be defined either broadly or more specifically; that is, learners can monitor their general performance on a whole performance test (global judgment; Bol et al., 2005) or focus on task solution for a very specific task or item (local judgments), for which they have access to task-specific cues (Frank & Kuhlmann, 2017; Händel, de Bruin, et al., 2020; Koriat, 1997).

So far, little is known regarding monitoring different cognitive objects of learning, that is, whether students

monitor different objects of their learning and how this monitoring relates to each other. The current study focuses on the different objects of monitoring processing and retrieval in different phases, as displayed in Table 1.

Aims and Hypotheses

The aim of the study was to investigate whether different objects of monitoring can be identified and how they relate to each other within and across preparation and testing phases. The study examined two major hypotheses regarding metacognitive monitoring.

First, we assume that metacognitive monitoring relates to different objects. We expect not only different monitoring objects related to different phases (namely, preparation and testing phase) but also related to processing and retrieval.

Hypothesis (H1): Metacognitive monitoring can be distinguished according to four different objects of monitoring: monitoring learning effectiveness, monitoring comprehension, monitoring test processing, and monitoring correctness of given answers.

Second, the relationship of monitoring different objects within and across different phases was investigated. We expect that monitoring objects are interrelated and that this applies especially when these objects refer to the same phase.

Hypothesis (H2): Monitoring different objects within one phase is stronger interrelated than monitoring different objects across phases.

Method

The current study was a cross-sectional study in the ecologically valid setting of a regular psychology course for undergraduate teacher training students. The main objective of the study was to investigate how students monitored their studying and their performance during test preparation as well as regarding their personal test processing. All study materials referred to the situation of test preparation or test processing and enabled a situation-specific assessment of students' metacognitive monitoring. Students answered several questionnaire items regarding monitoring strategies and provided item-specific monitoring judgments.

Procedure

The study took place within a course setting in a regular study term. The study referred to the lecture topic "empirical methods in education." This topic spanned three weekly lecture sessions at the beginning of the term. Subsequently, in the next (fourth) session, students had the opportunity to participate in an ungraded performance test and to answer corresponding questionnaire items. The performance test and survey questions were announced at the beginning of the lecture period and were administered via paper-and-pencil material during the lecture time. Students were informed that the test displays a learning opportunity to review previous lecture content and to become familiar with the test format of the final exam. It can thus be regarded as curricular valid. The performance test consisted of 14 multiple-choice items with one correct answer out of four answer possibilities. First, students reported to what extent they had monitored their learning in the preparation phase. Second, they reported to what extent they had monitored their comprehension of the lecture content via monitoring strategies. Third, students answered the performance test and provided monitoring judgments for each item of the performance test. Fourth, students reported on their implemented monitoring strategies during the test processing.

Sample

A power analysis was conducted to calculate the required sample size for the hypothesized factorial model including four separate monitoring objects. For a power of $(1-\beta) = .95$, a sample size of $N_p = 145$ was needed with an alpha error of .05 ($df = 224$), 23 manifest variables, and an expected effect size of RMSEA = .05. The participants were students of a psychology lecture for undergraduate students. All students enrolled in the course were invited to voluntarily participate in the study. They were not paid for participation. A total of 184 students took part in the study, and we did not exclude any students from participation; that is, the sample size of the recent study was sufficient to detect the statistical effects of interest and to test the hypotheses. Participants' privacy was protected, all data were anonymized, and students were not disadvantaged by nonparticipation. Informed consent of the participants was obtained by virtue of submission of the performance test and the respective questionnaire.

Most of the participants were first-year students (86.9 %). The majority of students were female (81.0 %), which is typical for university introductory courses in this

field of study. Students' GPA can be regarded as average, $M = 2.43$, $SD = 0.50$, on a scale ranging from 1 to 6 with lower values indicating better grades.

Little's MCAR (missing completely at random) test showed that missing data were missing completely at random, $\chi^2(1,119) = 1,106.874$, $p > .05$. The nonresponse missing rate ranged from 1% to 10% on the item level. Accordingly, the mice package by van Buuren and Groothuis-Oudshoorn (2011) was used with the predictive mean matching algorithm for multiple imputation of missing data. It created five imputed datasets for which pooled statistics based on a univariate combination of the imputed datasets were calculated. Accordingly, pooled means (M) as well as standard errors of the pooled means (SEM) are reported.

Instruments

The study investigated four aspects of monitoring – two aspects related to the preparation phase and two aspects related to the testing phase. Monitoring processing in the preparation and the testing phase was assessed via self-reported use of monitoring strategies regarding learning efficiency and test processing. Unfortunately, the study design did not allow for assessment of comprehension monitoring during the learning process via monitoring judgments as this would have required the assessment of spontaneous online judgments while the students engaged in learning (Jordano & Touron, 2018). Consequently, monitoring retrieval was assessed by self-reported use of comprehension monitoring strategies in the preparation phase. Monitoring correctness of the given answers in the performance test as a representation for monitoring retrieval in the testing phase was assessed by item-specific monitoring judgments.

The assessment of monitoring regarding the preparation phase included monitoring strategies related to learning effectiveness and monitoring strategies related to comprehension. Students answered the items on a 6-point Likert scale (1 = *not at all true* to 6 = *absolutely true*).

Monitoring learning effectiveness was assessed via three items adapted from the questionnaire scale "evaluation – monitoring" by McCardle and Hadwin (2015). A sample item was, "I checked to see if my approach was working." Students indicated to which degree they implemented strategies to monitor the usefulness of their learning approach.

Monitoring comprehension was assessed via a questionnaire scale with three items from the LIST inventory (Wild & Schiefele, 1994), which is a German analogue of the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1991). Students indicated to which degree

they implemented strategies to monitor their comprehension of course topics. A sample item was, "I asked myself questions about the material to make sure I understood everything."

The assessment of monitoring regarding the testing phase included monitoring strategies related to test processing and monitoring judgments.

Monitoring test processing was assessed via a newly developed and piloted questionnaire scale with three items focusing on whether students engaged in monitoring strategies regarding their test processing. Students indicated to which degree they implemented strategies to monitor their answer behavior in the previously solved performance test. A sample item was, "If I was unsure about a task, I thought through all the alternative answers again."

Monitoring correctness of given answers was assessed via 14 item-specific monitoring judgments in relation to performance test items (Händel, de Bruin, et al., 2020; Händel & Dresel, 2018). After completing each test item of the performance test, students indicated whether they thought their answer to the respective item was correct or not ("Do you think your answer is correct?"; 1 = *yes*, 0 = *no*). The judgments were printed in the same booklet as the test items, so that students had direct access to the test items and their respective answers when judging the correctness of their answers.

To measure the accuracy of students' judgments, bias was calculated. It reflects the degree of underconfidence (negative bias values) or overconfidence (positive bias values). Bias is computed as the signed difference between performance p_i and judgments c_i, averaged over the 14 performance test items (Schraw, 2009a). Values close to –1 point to high underconfidence, and values close to 1 indicate high overconfidence.

Data Analysis

For the empirical identification of different objects of students' monitoring (H1), three confirmatory factor analyses were computed. The hypothesized 4-factor model distinguished the four objects of monitoring: monitoring learning effectiveness, monitoring comprehension, monitoring test processing, and monitoring correctness of given answers. This model was compared with alternative models. The 2-factorial model distinguished the two factors monitoring in the preparation phase and monitoring in the testing phase, and the 1-factorial model comprised all objects of monitoring (see Table 2). We report latent correlations between the factors (monitoring objects) and used χ^2-differences to compare the fit of the nested models. As the bias score representing monitoring
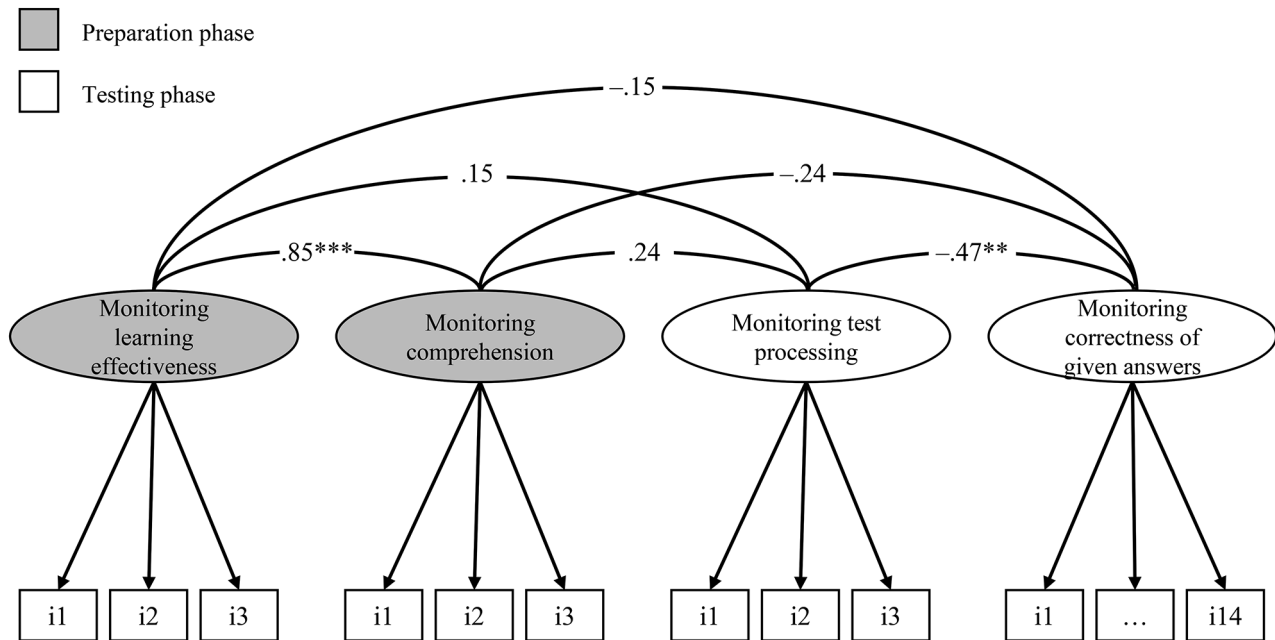
Note. **p < .01. ***p < .001.

Figure 1. Interrelations of the four latent factors representing four monitoring objects in the preparation phase (learning effectiveness and comprehension) and in the testing phase (test processing and correctness of given answers).

retrieval in the testing phase is a measure of accuracy (i.e., quality of monitoring) while monitoring the other objects are self-report data focusing on the quantity of strategy use rather than its quality, the factor analyses were additionally conducted using the 14 raw judgments instead of bias scores. The results also confirm the hypothesized 4-factor model (CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.61).

To examine the relationship between different objects of metacognitive monitoring within and across phases (H2), the latent correlations between the different objects were constrained, and the model fit results were compared via χ^2 tests. In the first model, latent correlation between monitoring objects within the preparation phase and correlations across the two phases were constrained to be equal. In the second model, latent correlation between monitoring objects within the testing phase and correlations across the two phases were constrained to be equal. If the constrained models lead to a significant drop in the model fit, this would indicate significantly different correlations.

Results

Different Objects of Monitoring: Confirmatory Factor Analyses (H1)

The confirmatory factor analyses showed that the model fit improved by distinguishing more factors (see Table 2). The hypothesized 4-factor model showed the best fit, which indicates that four different objects of students' monitoring could be empirically identified, namely, monitoring learning effectiveness, monitoring comprehension, monitoring test processing, and monitoring correctness of given answers (see Figure 1). The χ^2 difference tests comparing the models were each significant, one factor versus four factors: $\Delta\chi^2(6) = 161.940$, $p < .001$; two factors versus four factors: $\Delta\chi^2(5) = 47.590$, $p < .001$.

The 4-factor model distinguished between monitoring learning effectiveness ($M = 3.62$, $SEM = 0.06$), monitoring comprehension ($M = 3.90$, $SEM = 0.07$), monitoring test processing ($M = 4.63$, $SEM = 0.06$), and monitoring correctness of given answers indicated by bias ($M = -0.05$, $SEM = 0.03$). The composite reliabilities based on the factor loadings (Jöreskog, 1971) were acceptable for all scales, $.71 \leq \rho \leq .84$.

Table 2. Model fit indices for the hypothesized and alternative models of metacognitive monitoring

Model	χ^2	df	CFI	TLI	RMSEA	SRMR
Hypothesized model: 4 factors	241.080	224	.958	.952	.020	.065
Alternative model: 2 factors	288.670**	229	.852	.836	.038	.072
Alternative model: 1 factor	403.020***	230	.571	.528	.064	.089

Note. 1 factor = general factor for metacognitive monitoring; 2 factors = monitoring during the preparation phase vs. monitoring during the testing phase; 4 factors = monitoring learning effectiveness, monitoring comprehension, monitoring test processing, and monitoring correctness of given answers. **$p < .01$. ***$p < .001$.

Correlations Between Monitoring Objects Within and Across Phases (H2)

As can be seen in Figure 1, the two latent factors representing monitoring objects in the preparation phase (monitoring learning effectiveness and monitoring comprehension) were strongly interrelated. The more strategies students implemented to monitor their learning effectiveness, the more strategies they used to monitor their comprehension of the lecture content. The two latent factors representing monitoring objects in the testing phase (monitoring test processing and monitoring correctness of given answers indicated by bias) were interrelated to a medium-to-strong degree. The more strategies students implemented to monitor their test processing, the less overconfident they were in their given answers. Latent correlations between monitoring objects across preparation and testing phase were nonsignificant.

To test for differences between the latent correlations, two constrained models were calculated and compared with the hypothesized 4-factor model without constraints. The hypothesized 4-factor model showed a better model fit than the constrained model that forced latent correlations within the preparation phase and across phases to be equal, $\Delta\chi^2(2) = 36.932$, $p < .001$. The constrained model that forced latent correlations within the testing phase and across phases to be equal showed a similar model fit as the hypothesized model according to the chi-square test, $\Delta\chi^2(2) = 2.648$, $p > .05$. Still, as the hypothesized model shows an appropriate model fit, this indicates that different monitoring objects within the testing phase tend to be more strongly correlated than monitoring objects across different phases.

Discussion

The current study contributes with two main findings to the literature on metacognitive monitoring. First, metacognitive monitoring constitutes different monitoring processes. With a focus on monitoring cognitive aspects of learning (Winne & Hadwin, 1998), the current study found that monitoring learning effectiveness, monitoring comprehension, monitoring test processing, and monitoring correctness of given answers (indicated by bias as a score of monitoring judgment accuracy) can be distinguished. In the current study, monitoring not only referred to different objects, but these objects have relevance at different phases (e.g., monitoring learning behavior vs. monitoring testing behavior). This is in line with current experimental research that studied one specific monitoring activity, namely, explaining to (fictious) others, and indicated that timing matters (Lachner et al., 2020). Although the implemented test in the current study was a non-graded test, students seemed to take great care to monitor their test processing by monitoring strategies, as indicated by the degree to which students implemented monitoring strategies. Students reported descriptively higher scores regarding strategies to monitor test processing in the testing phase than regarding learning effectiveness and comprehension in the preparation phase. Accordingly, the current study reveals that students direct their attention to different objects of monitoring. Similarly, in a current video-based study on collaborative group work, monitoring was related to different objects in the preparation phase, and the ratings revealed different frequency of strategy use (e.g., most often to learning effectiveness; Malmberg et al., 2022). Future studies can build on this to encourage students to monitor specific objects. In addition, our study results suggest considering different objects of metacognitive monitoring not only in assessing monitoring but also when it comes to fostering monitoring. Second, the latent correlations showed that different monitoring objects within the preparation phase (monitoring learning effectiveness and monitoring comprehension) were more strongly related to each other than across phases. Students who monitored their learning effectiveness were very likely to monitor their comprehension and vice versa. The same applies to monitoring different objects within the testing phase (correlation between monitoring test processing and monitoring correctness of given answers indicated by bias). Students who monitored their test

processing were less overconfident regarding their given answers. However, metacognitive monitoring objects implemented across different phases were not significantly correlated to each other. Students seem to engage in different monitoring behaviors during preparation and testing. The results of the current study extend current research that showed only weak correlations between monitoring strategies and judgment accuracy (Dörrenbächer-Ulrich et al., 2021; Golke et al., 2022; Händel & Dresel, 2022). While Händel and Dresel (2022) also used a situation-specific approach, they investigated the correlation of comprehension monitoring in the preparation phase with the accuracy of monitoring judgments (i.e., monitoring correctness of given answers in the testing phase). Hence, the current study went one step further and considered different objects of monitoring within and across different phases. The lower overestimation of students who engaged in monitoring test processing represents an important finding as it shows that monitoring strategies and judgment bias are related to each other if they occur within the same phase.

The two main results of the study point to the relevance of the consideration of different objects within different phases of metacognitive monitoring (Kramarski & Michalsky, 2009; Nelson & Narens, 1990). The current research goes beyond the need for a situation-specific (in this case the situation of a concrete performance test) and subject-specific, fine granulated assessment of monitoring strategies (Rovers et al., 2019), and calls for an object-specific consideration of metacognitive monitoring (Azevedo, 2009; Dentakos et al., 2019).

Limitations and Future Directions

The situation-specific assessment of metacognitive monitoring related to a concrete performance test is a methodological advantage when investigating the relationship between the different objects of monitoring, assessed via monitoring strategies and monitoring judgments. Still, a potential limitation of this procedure is that the assessments of several monitoring aspects regarding one task could potentially influence each other as they function as an intervention (Händel, Harder, et al., 2020; Naujoks et al., 2022).

Significant and higher correlations were found between different objects of monitoring if the objects of monitoring referred to the same phase (either preparing or testing) compared to monitoring objects across different phases. However, monitoring judgments in the current study were restricted to the testing phase and the correlations between monitoring effectiveness and monitoring comprehension were quite high. Hence, the study does not provide any information on whether or how monitoring strategies and comprehension monitoring via monitoring judgments during the preparation phase are related to each other. While we used the typical procedure of assessing item-specific judgments in the testing phase, future research should consider (spontaneous) monitoring judgments during the preparation phase (Ariel et al., 2021; Jordano & Touron, 2018). This requires a more process-based/online assessment of metacognitive monitoring but has the potential to help understand the interrelations of monitoring strategies and monitoring judgments within and across different phases of learning. On the one hand, the use of monitoring strategies during test processing can lead to less overconfidence (Koriat et al., 2002). On the other hand, when students recognize that they judge many items as having being solved incorrectly, they might invest in further monitoring strategies.

Finally, a limitation of the study is that monitoring strategies were only assessed regarding their use, that is, to which degree students implemented monitoring strategies. By contrast, bias does not display the amount of monitoring judgments but its quality – that is, how accurate students are at judging their performance. Consequently, as a prospect for future research, studies should also consider quality aspects of monitoring strategies. In a current study with a situation-specific assessment, the quantity and quality of the use of monitoring strategies (in this case related to comprehension monitoring) were strongly correlated but were separate factors of metacognitive monitoring (Händel & Dresel, 2022). However, while the quality of the monitoring strategies used was again based on a self-report, other measures such as micro-analyses (Dörrenbächer-Ulrich et al., 2021) might provide further insights.

Overall, the current results provide room for further investigation, especially with regard to regulation of studying and test processing. More specifically, the question to be addressed is whether students adapt their answers according to the monitoring strategies used during testing (e.g., change their answers, less often opt-out answers; Couchman et al., 2016; Papanastasiou & Stylianou-Georgiou, 2022) or whether they adapt their metacognitive judgments with respect to their monitoring strategies.

References

Alexander, P. A., Graham, S., & Harris, K. R. (1998). A perspective on strategy research: Progress and prospects. *Educational Psychology Review*, *10*(2), 129–154. https://doi.org/10.1023/a:1022185502996

Ariel, R., Karpicke, J. D., Witherby, A. E., & Tauber, S. K. (2021). Do judgments of learning directly enhance learning of educational materials? *Educational Psychology Review*, *33*(2), 693–712. https://doi.org/10.1007/s10648-020-09556-8

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). Motivated learning: Studying student * situation transactional units. *European Journal of Psychology of Education*, *14*(1), 41–55. https://doi.org/10.1007/bf03173110

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education*, *73*(4), 269–290. https://doi.org/10.3200/JEXE.73.4.269-290

Broadbent, J., Panadero, E., Lodge, J. M., & Fuller-Tyszkiewicz, M. (2022). The self-regulation for learning online (SRL-O) questionnaire. *Metacognition and Learning, 18.* 135–163. https://doi.org/10.1007/s11409-022-09319-6

Couchman, J. J., Miller, N. E., Zmuda, S. J., Feather, K., & Schwartzmeyer, T. (2016). The instinct fallacy: The metacognition of answering and revising during college exams. *Metacognition and Learning*, *11*(2), 171–185. https://doi.org/10.1007/s11409-015-9140-8

Craig, K., Hale, D., Grainger, C., & Stewart, M. E. (2020). Evaluating metacognitive self-reports: Systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning*, *15*(2), 155–213. https://doi.org/10.1007/s11409-020-09222-y

Credé, M., & Phillips, L. A. (2011). A meta-analytic review of the Motivated Strategies for Learning Questionnaire. *Learning and Individual Differences*, *21*(4), 337–346. https://doi.org/10.1016/j.lindif.2011.03.002

de Bruin, A. B. H., Kok, E. M., Lobbestael, J., & de Grip, A. (2016). The impact of an online tool for monitoring and regulating learning at university: Overconfidence, learning strategy, and personality. *Metacognition and Learning*, *12*(1), 21–43. https://doi.org/10.1007/s11409-016-9159-5

Dentakos, S., Saoud, W., Ackerman, R., & Toplak, M. E. (2019). Does domain matter? Monitoring accuracy across domains. *Metacognition and Learning*, *14*(3), 413–436. https://doi.org/10.1007/s11409-019-09198-4

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

Dresel, M., Schmitz, B., Schober, B., Spiel, C., Ziegler, A., Engelschalk, T., Jöstl, G., Klug, J., Roth, A., Wimmer, B., & Steuer, G. (2015). Competencies for successful self-regulated learning in higher education: Structural model and indications drawn from expert interviews. *Studies in Higher Education*, *40*(3), 454–470. https://doi.org/10.1080/03075079.2015.1004236

Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology*, *43*(5), 680–693. https://doi.org/10.1037/xlm0000332

Golke, S., Steininger, T., & Wittwer, J. (2022). What makes learners overestimate their text comprehension? The impact of learner characteristics on judgment bias. *Educational Psychology Review*, *43*(4), 2405–2450. https://doi.org/10.1007/s10648-022-09687-0

Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research*, *77*(3), 334–372. https://doi.org/10.3102/003465430303953

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. https://doi.org/10.1007/978-1-4419-5546-3_2

Hacker, D. J., & Bol, L. (2019). Calibration and self-regulated learning. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 647–677). https://doi.org/10.1017/9781108235631.026

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*(1), 160–170. https://doi.org/10.1037//0022-0663.92.1.160

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 429–455). Psychology Press.

Händel, M., de Bruin, A. B. H., & 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

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*, Article 102229. https://doi.org/10.1016/j.lindif.2022.102229

Händel, M., Harder, B., & Dresel, M. (2020). Enhanced monitoring accuracy and test performance: Incremental effects of judgment training over and above repeated testing. *Learning and Instruction*, *65*, Article 101245. https://doi.org/10.1016/j.learninstruc.2019.101245

Jordano, M. L., & Touron, D. R. (2018). How often are thoughts metacognitive? Findings from research on self-regulated learning, think-aloud protocols, and mind-wandering. *Psychonomic Bulletin & Review*, *25*(4), 1269–1286. https://doi.org/10.3758/s13423-018-1490-1

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*(2), 109–133. https://doi.org/10.1007/bf02291393

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. (2019). Confidence judgments: The monitoring of object-level and same-level performance. *Metacognition and Learning*, *14*, 463–478. https://doi.org/10.1007/s11409-019-09195-7

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

Kramarski, B., & Michalsky, T. (2009). Three metacognitive approaches to training pre-service teachers in different learning phases of technological pedagogical content knowledge. *Educational Research and Evaluation*, *15*(5), 465–485. https://doi.org/10.1080/13803610903444550

Lachner, A., Backfisch, I., Hoogerheide, V., van Gog, T., & Renkl, A. (2020). Timing matters! Explaining between study phases enhances students' learning. *Journal of Educational Psychology*, *112*(4), 841–853. https://doi.org/10.1037/edu0000396

Lan, W. (2005). Self-monitoring and its relationship with educational level and task importance. *Educational Psychology*, *25*(1), 109–127. https://doi.org/10.1080/0144341042000294921

Law, M. K. H., Stankov, L., & Kleitman, S. (2022). I choose to opt-out of answering: Individual differences in giving up behaviour on cognitive tests. *Journal of Intelligence*, *10*(4), Article 86. https://doi.org/10.3390/jintelligence10040086

Maki, R. H., & McGuire, M. J. (2002). Metacognition for text: Findings and implications for education. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 39–67). Cambridge University Press.

Malmberg, J., Saqr, M., Järvenoja, H., & Järvelä, S. (2022). How the monitoring events of individual students are associated with phases of regulation. *Journal of Learning Analytics*, *9*(1), 77–92. https://doi.org/10.18608/jla.2022.7429

McCardle, L., & Hadwin, A. F. (2015). Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. *Metacognition and Learning*, *10*, 43–75. https://doi.org/10.1007/s11409-014-9132-0

Miele, D. B., & Scholer, A. A. (2017). The role of metamotivational monitoring in motivation regulation. *Educational Psychologist*, *53*(1), 1–21. https://doi.org/10.1080/00461520.2017.1371601

Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning*, *6*, 303–314. https://doi.org/10.1007/s11409-011-9083-7

Naujoks, N., Harder, B., & Händel, M. (2022). Testing pays off twice: Potentials of practice tests and feedback regarding exam performance and judgment accuracy. *Metacognition and Learning*, *17*(2), 479–498. https://doi.org/10.1007/s11409-022-09295-x

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Papanastasiou, E. C., & Stylianou-Georgiou, A. (2022). Should they change their answers or not? Modelling achievement through a metacognitive lens. *Assessment in Education: Principles, Policy & Practice*, *29*(1), 77–94. https://doi.org/10.1080/0969594x.2022.2053945

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, *31*(6), 459–470. https://doi.org/10.1016/S0883-0355(99)00015-4

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. University of Michigan.

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/s11409-019-09188-6

Schraw, G. (2009a). 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

Schraw, G. (2009b). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 415–429). Routledge.

Schuster, C., Stebner, F., Geukes, S., Jansen, M., Leutner, D., & Wirth, J. (2023). The effects of direct and indirect training in metacognitive learning strategies on near and far transfer in self-regulated learning. *Learning and Instruction*, *83*, Article 101708. https://doi.org/10.1016/j.learninstruc.2022.101708

Sobocinski, M., Järvelä, S., Malmberg, J., Dindar, M., Isosalo, A., & Noponen, K. (2020). How does monitoring set the stage for adaptive regulation or maladaptive behavior in collaborative learning? *Metacognition and Learning*, *15*(2), 99–127. https://doi.org/10.1007/s11409-020-09224-w

Stebner, F., Schuster, C., Weber, X.-L., Greiff, S., Leutner, D., & Wirth, J. (2022). Transfer of metacognitive skills in self-regulated learning: Effects on strategy application and content knowledge acquisition. *Metacognition and Learning*, *17*(3), 715–744. https://doi.org/10.1007/s11409-022-09322-x

Stylianou-Georgiou, A., & Papanastasiou, E. C. (2017). Answer changing in testing situations: The role of metacognition in deciding which answers to review. *Educational Research and Evaluation*, *23*(3/4), 102–118. https://doi.org/10.1080/13803611.2017.1390479

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Wild, K. P., & Schiefele, U. (1994). Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens [Learning strategies of university students: Factor structure and reliability of a new questionnaire]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *15*(4), 185–200.

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

Winne, P. H., & Muis, K. R. (2011). Statistical estimates of learners' judgments about knowledge in calibration of achievement. *Metacognition Learning*, *6*(2), 179–193. https://doi.org/10.1007/s11409-011-9074-8

Wirth, J., & Leutner, D. (2008). Self-regulated learning as a competence. *Journal of Psychology*, *216*(2), 102–110. https://doi.org/10.1027/0044-3409.216.2.102

History

Funding

ORCID

Marion Händel
 https://orcid.org/0000-0002-3069-5582
Nick Naujoks-Schober
 https://orcid.org/0000-0003-3211-6887
Markus Dresel
 https://orcid.org/0000-0002-2131-3749

Dr. Nick Naujoks-Schober
Institute of Psychology
Friedrich-Alexander University Erlangen-Nuremberg
Regensburger Straße 160
90478 Nürnberg
Germany
nick.naujoks@fau.de