Individual differences in local and global metacognitive judgments

Marion Händel[1] · Anique B. H. de Bruin[2] · Markus Dresel[3]

Abstract

Extensive research has been conducted to understand how accurately students monitor their studying and performance via metacognitive judgments. Moreover, the bases of students' metacognitive judgments are of interest. While previous results are quite consistent regarding the importance of performance for the accuracy of metacognitive judgments, results regarding motivational and personality variables are rather heterogeneous. This paper reports on two studies that simultaneously examined the predictive power of several performance, motivational, and personality variables on metacognitive judgments. The studies investigated a set of judgments (local and global postdictions in Study 1 and global pre- and postdictions in Study 2) and accuracy scores (bias, sensitivity, and specificity) in two different settings. Individual differences in judgments and judgment accuracy were studied via hierarchical regression analyses. Study 1 with $N = 245$ undergraduate students identified performance and domain-specific self-concept as relevant predictors for judgments after test taking. This was consistently found for local and global judgments. Study 2 with $N = 138$ undergraduate students hence focused on domain-specific self-concept and extended results to predictions. Study 2 replicated results for global postdictions but not predictions. Specifically, before task processing, students' judgments relied mostly on domain-specific self-concept but not on test performance itself. The studies indicate that different judgments and measures of judgment accuracy are needed to obtain comprehensive insights into individual differences in metacognitive monitoring.

✉ Marion Händel
marion.haendel@fau.de

1   Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

2   Maastricht University, Maastricht, The Netherlands

3   University of Augsburg, Augsburg, Germany

In a perfect (self-regulated learning) world, learners consistently monitor their progress and adapt their learning activities according to the results of these monitoring judgments (Nelson and Narens 1990). Based on their metacognitive monitoring judgments, students are supposed to initiate regulation processes. Therefore, monitoring activities should affect students' future effort and learning behavior and should consequently lead to higher performance (Winne and Hadwin 1998). However, this interplay of monitoring and regulation works smoothly only if monitoring judgments are accurate. Indeed, students are predominantly inaccurate judges of their individual performance and tend to overestimate their test performance (e.g., Händel and Dresel 2018; Kruger and Dunning 1999). Hence, the question arises as to why students provide biased judgments or, in other words, what the bases of students' (inaccurate) judgments are. The current two studies used an individual differences approach to understand why some students are able to accurately judge their performance and others are not. Based on theoretical and empirical grounds, we simultaneously studied performance, motivation, and personality as possible predictors for metacognitive judgments.

Theoretical background

Metacognitive judgments are inferential in nature (Perfect and Schwartz 2002). These inferences are based on cues people have access to when judging performance. The cue-utilization approach by Koriat (1997) provides a theoretical model suitable to understand different causes for metacognitive judgments and their accuracy. According to this approach, the accuracy of judgments depends on the availability of specific cues. The cues students use to generate their judgments can be based on previously gathered information—information-based (or theory-based) cues. In this case, students might base their judgments on preconceived notions about competence, for example, due to their self-concept beliefs, prior success in the respective domain, or time and effort for studying. In addition, students might base their judgments on their concrete experiences during task processing, like experienced retrieval fluency or detection of potential difficulties with the tasks—experience-based cues (Koriat et al. 2008).

Metacognitive judgments

To understand how these two types of cues affect judgment accuracy, one needs to distinguish between different types of judgments (for an overview, see Nelson and Narens 1990; Schraw 2009). First, judgments of test performance can differ according to their grain size; that is, they can be made on a local or global level. While global judgments are assessed at the test level, local judgments refer to each single task. In addition, judgments are distinguished according to the point in time at which they occur: Predictions are made before test taking and postdictions are made after test taking. Research on metacognitive judgments offers a wide variety of accuracy scores for local as well as global judgments. The scores can be distinguished into absolute accuracy (synonymous to calibration) and relative accuracy scores (synonymous to discrimination; Bol and Hacker 2012). Absolute accuracy scores measure the overall precision of judgments compared to performance, while relative accuracy scores measure the relationship of judgments and performance scores. Global judgments allow calculating their over- or underconfidence (bias score). Assessing local judgments in a dichotomous rather than a continuous way furthermore offers the opportunity to analyze their accuracy via several relative accuracy scores (Rutherford 2017; Schraw et al. 2013). Overall, it is suggested to

measure accuracy by more than only one score. Schraw et al. (2013) favor sensitivity and specificity as these two scores consider items to be answered either correctly or incorrectly. In our studies, we calculated them in addition to the bias score.

Earlier work distinguishing between information-based and experience-based cues usually referred to local judgments (Koriat et al. 2008). Local judgments can be based on information gathered before task completion (information-based cues like perceptions of one's own exam preparation) as well as on experiences during task processing (e.g., experienced task difficulty). Both experience- and information-based cues can display relevant sources for performance judgments but, of course, might differ in their validity. Compared to global judgments, local judgments can more easily be based on experience-based (task-specific) cues because students directly experience retrieval fluency or potential difficulties with the tasks. Global judgments might be driven by cues such as domain-familiarity or self-concept (information-based cues) because task-specific cues are not available or would need to be averaged across all test items (Dunlosky and Lipko 2007; Gigerenzer et al. 1991). However, considering global judgments, judgments made before or after test taking need to be distinguished. Postdictions can be based on both information- and experienced-based cues. For example, a student might have information about his or her previous performance in the domain (information-based cue) and might have experienced fast or slow retrieval (experience-based cue; Koriat et al. 2008; Thiede et al. 2010). When making predictions, students have not had any prior experiences with the concrete task. Hence, information-based cues like self-concept are of higher relevance for predictions. Higher accuracy of postdictions in contrast to predictions, which were usually found in the literature, might be explained via access to more relevant cues (Pierce and Smith 2001).

Individual differences in metacognitive judgments

In addition to experienced-based cues generated during task completion, several individual factors might provide information based-cues related to the monitoring of personal understanding or performance. That is, on the one hand, students might make use of their experiences during testing to make their judgments. Retrieval fluency or application of prior knowledge for task processing might, for example, serve as relevant experiences (Dinsmore and Parkinson 2013; Händel and Dresel 2018). On the other hand, motivational and personality variables may facilitate the use of information-based cues like perceptions about prior success or personal exam preparation. Currently, motivational and personality variables, such as self-concept, the Big Five personality factors, or narcissism, are in the focus of research on individual differences (Buratti et al. 2013; Kröner and Biermann 2007; Kruger and Dunning 1999; Pallier et al. 2002; Schaefer et al. 2004). It is assumed that learner characteristics like optimism or narcissism lead to general tendencies to (over-)estimate personal performance (de Bruin et al. 2017). Hence, individuals with positive feelings and expectancies are supposed to judge their achievements more favorable. Students' judgments might also be driven by motivational or judgment tendencies, which might result in stable judgments over time, as indicated by previous research (e.g., Hacker et al. 2000). In a study by Foster et al. (2017), students remained overconfident in their global judgments even after multiple exam predictions. However, training studies focusing on local judgments indicate that students enhanced their judgment accuracy after being informed about their importance or after receiving feedback on their accuracy (Händel et al. 2019; Roelle et al. 2017). These differences in the stability of local and global judgments over time point to differences in cues that are available

and used when providing performance judgments. Hence, global judgments might relate more strongly to (rather stable) information-based cues while local judgments might be more easily based on experiences with the concrete task (i.e. experience-based cues). In the following, we discuss potential factors influencing (the accuracy of) metacognitive judgments. We provide respective empirical results that influenced our assumptions regarding which factors should contribute to which type of judgments or judgment accuracy.

Individual test performance Probably the most prominent influencing factor on metacognitive judgments is individual test performance. In general, test performance positively correlates with judgments (higher performance relates to higher judgments) and negatively correlates with overconfidence (the higher the performance the lower the overestimation). This pattern of results is called the unskilled and unaware effect, and it has appeared across several (classroom) studies for different domains and types of judgments (Bol and Hacker 2001; Kruger and Dunning 1999; Nietfeld et al. 2005). It is supposed that differences in performance influence the use of available (experience-based) cues in order to realize individual strengths and weaknesses (Kornell and Hausman 2017; Kruger and Dunning 1999).

Usually, studies on the unskilled and unaware effect split students into performance quartiles. This revealed significant and strong differences in judgments between quartiles (e.g., Händel and Fritzsche 2016; Miller and Geraci 2011b). This prevailing procedure, however, possibly inherits artifacts, for example because high performing students—by definition—cannot overestimate their performance as much as low performing students, and vice versa (better-than-average heuristic; Krueger and Mueller 2002). Correlational or regression approaches offer the advantage to consider performance as a continuous variable and avoid composing artificial performance groups; see, for example, the study by Serra and DeMarree (2016). In this study, individual test performance significantly predicted the global postdictions of final course exams with small to moderate beta-weights. Similarly, Händel and Fritzsche (2015) reported significant and positive correlations between test performance and judgment accuracy. When students provide local judgments, performance should have a higher impact than for global judgments because students then judge the individual performance of tasks just processed beforehand. In line with this, test performance as an experience-based cue should be relevant for judgments made after but not before testing. In addition to current test performance, students seem to base their judgments on their prior knowledge, as studies with open responses have indicated (Dinsmore and Parkinson 2013; Händel and Dresel 2018). In line with effects of current test performance, higher prior performance (suggesting a more elaborate knowledge base) should lead to higher and more accurate judgments.

Gender Although not the primary focus of previous studies, some of them controlled for gender effects on judgment accuracy when investigating individual differences in metacognitive judgments. For example, in studies with general knowledge questions or questions implemented in a college course, female students provided lower and less overestimated confidence scores than male students (de Bruin et al. 2017; Buratti et al. 2013). However, overall, findings seem mixed (cf. Jonsson and Allwood 2003). In addition, gender differences might be domain-specific and relate to gender differences in self-efficacy or self-concept as information-based cues (Jansen et al. 2014; Pallier 2003; Retelsdorf et al. 2015). For example, in mathematics as a stereotypical male-dominated domain, male students scored higher in self-efficacy, self-concept, and local as well as global judgments (Morony et al. 2013). One current study focused on gender differences in local and global judgments

and their accuracy regarding spatial ability tasks. It resulted that female student scored lower on each measure but were not less accurate than male students (Ariel et al. 2018). The authors discussed that male and female students possibly not only use different strategies to solve the spatial tasks but also different cues to judge their performance.

The big five personality factors According to the five factor model, the five personality factors are (1) openness, (2) conscientiousness, (3) extraversion, (4) agreeableness, and (5) neuroticism (McCrae and Costa 1999). Theoretically, some relationships between Big Five factors and judgments can be assumed. Namely, neuroticism is associated with negative feelings and described by adjectives like worrying (McCrae and Costa 1987; Williams 1992). Neuroticism might be associated with lower expectations about personal success and thereby with lower judgments. In contrast, extraversion, conscientiousness, and agreeableness having positive correlations with optimism (Sharpe et al. 2011) are supposed to lead to higher judgments. While extraverted people are assumed to be overconfident in the consequence of their, on average, more positive self-esteem, conscientious or agreeable people might be concerned about projecting a false sense of confidence and thereby not necessarily be overconfident (Agler et al. 2019).

Empirical evidence provides support for positive relations of monitoring judgments on the one hand (mainly local judgments) and extraversion (Buratti et al. 2013; Schaefer et al. 2004), conscientiousness (Schaefer et al. 2004), and openness on the other hand (Agler et al. 2019; Dahl et al. 2010; Schaefer et al. 2004). Negative relations were found for neuroticism (Burns et al. 2016; Dahl et al. 2010). In detail, in a sample of undergraduate students, significant but small correlations of extraversion, conscientiousness, and openness with confidence and/or bias occurred (Schaefer et al. 2004). That is, students with higher values in extraversion, conscientiousness, and openness provided higher and more overconfident local judgments. Another study with local judgments regarding general knowledge questions used hierarchical regressions to investigate the influence of the Big Five (Buratti et al. 2013). Controlling for age, gender and performance, two of the Big Five scales, namely openness and a combined scale with items about extraversion and narcissism, led to higher judgments. Again, the effects were quite small. In contrast, in a study with local judgments regarding intelligence tests, only neuroticism, and mostly for younger adults, was a significant and negative predictor for judgment level (Burns et al. 2016). Finally, a study with older adults found significant but weak influences of neuroticism (negatively related) and openness (positively related) on local judgments regarding general knowledge questions (Dahl et al. 2010). Overall, only some (but not consistently the same) of the Big Five scales were found to have effects on judgments. The reported effects were of a similar small effect size. However, a shortcoming of previous studies is that they investigated a subset of person characteristics and mostly did not control for other variables known to be relevant predictors for metacognitive judgment accuracy. Ideally, a comprehensive set of possible (cognitive, personal, and motivational) predictors is investigated simultaneously to reveal the bases of metacognitive judgments and their accuracy.

Narcissism Narcissism is characterized by an inflated sense of self and feelings of superiority (Küfner et al. 2015; Paulhus and Williams 2002). These feelings might promote the consideration of cues that lead to overestimation, as indicated in studies with managers and stock market investors (John and Robins 1994). A couple of studies investigated whether narcissism influences metacognitive judgments in academic contexts (e.g., Farwell and Wohlwend-Lloyd 1998). In two series of studies with local judgments, it was shown that narcissism moderately

predicted overconfidence on general knowledge questions (Campbell et al. 2004; Macenczak et al. 2016). A study that was implemented in a college course, however, found that narcissism contributed only to overconfidence of exam score predictions but not of task-specific judgments (de Bruin et al. 2017), which might be due to a larger influence of experience-based cues when making local judgments.

Optimism Optimism is a trait factor that describes individuals' positive expectancies about their future (Carver et al. 2010). These general expectancies might transfer to judgments of test performance. Hence, more optimistic people are assumed to provide higher judgments about their personal performance. However, their judgments do not necessarily need to be more accurate: When people are overoptimistic, they may provide overconfident judgments. So far, only one study has investigated optimism in an educational setting as a predictor for judgments but did not find significant relationships, neither for global nor for local judgment accuracy (de Bruin et al. 2017).

Self-concept and self-efficacy Self-concept refers to a cognitive representation of one's own abilities. It is described as a multifaceted construct (Marsh 1986, 1990; Marsh and Shavelson 1985). General self-concept is divided into non-academic and academic self-concept. An individual might hold different self-concepts for different domains. For example, a student might hold a positive self-concept in English language but a lower mathematical self-concept. Self-efficacy refers to expectations about successful task completion or achievement of goals. These beliefs are based on self-evaluations of one's skills (Bandura 1982). Self-efficacy is competence-based, prospective, and action-related (Luszczynska et al. 2005) and can be task-specific or general. Self-concept and self-efficacy both inherit a self-evaluation, and thus are interrelated in many respects; for differences between self-concept and self-efficacy, see Pajares and Miller (1994).

Higher scores in self-concept or self-efficacy should be correlated with higher metacognitive judgments as all three constructs are based on judgments of personal competence (Ackerman et al. 2002; Lösch et al. 2017; Stankov et al. 2012). There is some empirical evidence on the relevance of self-concept for judgments. However, as with many other potential variables, self-concept was studied without controlling for other potential variables (e.g. Efklides and Tsiora 2002; Jiang and Kleitman 2015). The study by Kröner and Biermann (2007) postulated and empirically proved a model of response confidence considering competence, self-concept, and judgments. In their study, undergraduate students provided local judgments regarding several performance tests and filled in a questionnaire regarding academic self-concept. Structural equation modeling indicated that a considerable amount of variance in the confidence factor was explained by the self-concept factor that was over and above the influence of test performance. In a study with fourth graders, correlations of a medium size were found between local confidence judgments in a spelling test and domain-specific self-concept (Fritzsche et al. 2012). Similarly, in another study with fourth graders, medium and positive correlations were found for reading self-concept and confidence ratings in a reading comprehension test (Kasperski and Katzir 2013). Finally, domain-specific self-concept and self-efficacy were significantly and moderately to strongly related with local confidence ratings in the domain of mathematics, as indicated by a study with adolescents of different countries (Morony et al. 2013). In sum, results indicate that self-concept, which represents information about prior success, may be a relevant predictor for all types of metacognitive judgments.

Aims of the current studies

So far, a fair amount of studies have investigated individual differences in metacognitive judgments. Different performance, personality, and motivational factors have been studied to understand how students generate metacognitive judgments. While performance could be shown as a relevant factor, motivational and personality factors in comparison seem less strongly related to confidence judgments and their accuracy. However, studies usually considered only some selected factors, did not study them simultaneously, focused only on bias as one of many metacognitive accuracy scores, and mostly did not control for the relationship with already known predictors like, for example, performance. The study of single factors or combinations of factors was an important first step but does little to get a comprehensive picture of relevant predictors. Hence, little is known about which of the individual factors are more relevant for accurate judgments than others.

Previous research on individual differences in judgment accuracy has been conducted mostly with local judgments. However, the accuracy of those judgments has been assessed often only via aggregated measures like bias or absolute accuracy, which partly contradicts the idea of local judgments. That is, to understand how specific learner characteristics are related to judgment accuracy, the implementation of further calibration measures is needed. For example, students scoring high on narcissistic, optimistic, or self-concept scales might be better in detecting items answered correctly (high sensitivity) but might avoid acknowledging if they were not able to answer some items (low specificity; Schraw et al. 2013). With regard to the grain size of judgments, previous research has focused on local judgments compared to global judgments. This is surprising, especially as personality factors are supposed to be more relevant for global judgments than for local judgments. However, this assumption has rarely been tested.

To overcome the limitations of previous studies, the current work simultaneously considers performance, personality, and motivational variables to clarify which of these variables are relevant when students provide local and global judgments. We conducted two studies to investigate which variables are related to judgments. We differentiated between local and global judgments, pre- and postdictions, and the accuracy measures bias, sensitivity, and specificity.

Research questions and hypotheses

We aimed to investigate which factors significantly contribute to pre- and postdicted local and global judgments and their respective accuracy scores. In general, we assumed that cognitive, motivational, and personality characteristics contribute to the level of the judgment and its accuracy.

> H1: Judgments and judgment accuracy are related to performance, motivation, and personality.

Considering the effect sizes of previous studies, we assumed actual performance to be a strong predictor. Conceptually, confidence judgments should be strongly related to other types of self-assessments like self-concept or self-efficacy. Hence, we expected that above and beyond

performance, perceived performance (domain-specific self-concept) is a significant predictor for judgments. Finally, due to rather weak and inconsistent findings regarding personality variables, we expected only small effects here.

Next, we considered differences in local and global postdictions. They both can be based on test performance. In contrast, local judgments are assumed to be based on task-specific cues rather than on trait-like variables. Hence, we propose:

> H2: Motivation and personality should be more strongly related to global postdicted judgments and accuracy scores than to local judgments and accuracy scores.

Finally, we compared the results for global pre- and postdictions. As task characteristics can provide helpful cues for students' postdictions but not their predictions, we made the following assumption:

> H3: Global predictions are more strongly related to self-concept beliefs but less strongly related to performance than global postdictions.

General method

Two studies were conducted to investigate possible predictors for different types of judgments and accuracy scores. Both studies are non-experimental and correlational studies with under-graduate teacher education students. Study 1 investigated a broad set of possible predictors for local and global postdictions regarding a mathematics test in a laboratory setting. In addition, all potential predictors were studied with regard to their relationship with global judgments in a natural exam setting in the domain of psychology. Study 2 focused on general as well as domain-specific self-concept and their relationship with global pre- and postdictions in an exam setting.

Study 1

Design of the Study

The study took place in regular educational psychology courses at two German universities. In the middle of the term, students voluntarily participated in a testing session. Students filled in a test regarding mathematical functions and provided local as well as global judgments. Afterwards, students answered a questionnaire regarding several motivational and personality variables. Students furthermore participated in a written course exam at the end of the term. The students were asked to judge their exam performance after having completed the exam (global postdiction).

Sample

Of 284 undergraduate students, 245 students participated in a testing session comprising a mathematics test and a questionnaire.[1] A subsample of $n = 128$ participants enrolled in a

---

[1] Data from students who filled in only the questionnaire ($n = 8$), only the mathematics test ($n = 16$), or who did not provide judgments ($n = 14$) were excluded from the analyses.

psychology course participated in a real exam, and provided a global judgment regarding this exam.[2] The majority of students were female (75.7%), a gender distribution typical for German teacher education programs (Schmidt 2013). Most students were enrolled in their first (73%) or second (18%) study year.

Instruments

Prior performance Students self-reported their high school grade (overall as well as specific to mathematics) as an indicator for prior performance. Grades were recoded as ranging from 1 (lowest degree) to 6 (highest degree).

Domain-specific performance tests Two performance tests were implemented in the study—for an overview, see Table 1. First, an extracurricular test in the domain of mathematics and, second, a real exam in the domain of psychology were used. Both tests were multiple choice tests. The mathematical test consisted of items about mathematical functions of a mathematics test for high school graduates and freshmen (MTAS; Lienert and Hofer 1972). To process the specific items, students needed to apply mathematical knowledge they should have acquired in high school. The psychology exam as a second performance criterion was developed by the first author. It was a curricularly valid test about the content of the according lecture. For both tests, several test forms were implemented in order to prevent students from copying their peers' work. Each test form contained the same items at different item positions. Within each performance test, items were coded as either 0 for incorrect or 1 for correct answers. For each performance test, mean scores ranging from 0 to 1 were calculated. Both tests revealed acceptable internal consistency (see Table 1). Students' performance did not significantly differ between the two versions in the mathematical test ($p = .90$), or between the four versions in the final psychology exam ($p = .66$).

Global performance judgments Both tests were followed by a global performance judgment. After completing the tests, the students were instructed to indicate the number of items they thought they had answered correctly. For both tests, the global judgments were recoded on the same scale as the performance tests.

Local judgments After each test item in the extracurricular test in mathematics, students were asked to provide a judgment about the correctness of their answer („Do you think you picked the correct answer?"). Students had to tick one of two boxes labeled with "yes" or "no," recoded into 1 for yes and 0 for no.[3] This dichotomous assessment of judgments seems a quite appropriate and quick procedure when assessing judgments regarding single choice items. That is, students knew that each item was scored as correct or incorrect and they could base their judgments on the same metric.

Motivational and personality variables We assessed three motivational constructs. Domain-specific self-concept was assessed via a 4-item scale (sample item: "Mathematics is not for

---

[2] Students who did not voluntarily provide the judgment in the exam situation ($n = 16$) were excluded from the study.

[3] Judgments regarding non-answered test items were recoded as missing because no metacognitive ability is required to judge an item as incorrectly answered if no answer to the test item has been given.

Table 1 Implemented measures in Study 1 and Study 2

Test	Topic	Item type	# items	α	Judgments
Study 1					
Mathematics test	Mathematical functions (e.g., derivation, logarithm)	Multiple choice (1 out of 3)	18	.73	Local + global
Psychology exam	Research methods, diagnostics and learning	Multiple choice (1 out of 4)	32	.69	Global
Study 2					
Psychology exam	Research methods, diagnostics and learning	Multiple choice (1 out of 4)	32	.75	Global

me", Cronbach's $\alpha = .89$) by Köller et al. (2000). Academic self-concept was assessed with a scale by Dickhäuser et al. (2002) with 5 items (sample item: "I am highly gifted at studying", $\alpha = .81$). Self-efficacy was assessed via a 3-item scale (sample item: "Usually, I can even solve hard and complex tasks", $\alpha = .78$) by Beierlein et al. (2012). Furthermore, optimism was assessed with an 8-item scale (sample item: "I think my studies will develop well", $\alpha = .86$), constructed by Satow and Schwarzer (1999). Narcissism was assessed with a 3-item scale (sample item: "I tend to strive for reputation or status", $\alpha = .76$), constructed by Küfner et al. (2015). The Big Five personality factors were assessed with a 21-item instrument by Rammstedt and John (2005), comprising the scales extraversion (sample item: "I let myself go, I am a social person", 4 items, $\alpha = .77$), agreeableness (sample item: "I give credit to others, I believe in the good in people", 4 items, $\alpha = .65$), conscientiousness (sample item: "I manage my tasks properly", 4 items, $\alpha = .73$), neuroticism (sample item: "I am easily depressed", 4 items, $\alpha = .76$), and openness (sample item: "I am open minded", 5 items, $\alpha = .78$). All items are answered on a 6-point Likert scale, ranging from 1 (*not at all true*) to 6 (*completely true*) with higher scores indicating higher agreement.

Data analyses

To account for the richness of information that local judgments provide, we calculated relative measures in addition to absolute measures (Bol and Hacker 2012). Bias scores were calculated for local judgments in the mathematics test as well as for global judgments in both domains. In addition, based on the 2×2 contingency table of task performance and monitoring judgment (see Table 2), sensitivity and specificity were calculated as relative measures (Rutherford 2017; Schraw et al. 2013).

Bias indicates the degree of underconfidence (negative bias values) or overconfidence (positive bias values) and is computed as the signed difference between performance p_i and judgments c_i, averaged over the n items (Formula 1). Bias ranges between -1 (underconfidence) and 1 (overconfidence). It provides information on the direction of overconfidence or underconfidence and was calculated for global as well as local judgments.

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

Sensitivity (see Formula 2—for indices, see Table 2) indicates the amount of accurately detected correct answers. Specificity (see Formula 3—for indices, see Table 2) indicates the

Table 2 2 × 2 Table of dichotomous local judgments

Judgment	Task performance	
	correct	incorrect
yes	a Hit/true positive	b False alarm/false positive
no	c Miss/false negative	d Correct rejection/true negative

amount of accurately detected incorrect answers. Sensitivity is calculated as the proportion of hits/true positives as measured by all items answered correctly. Specificity is measured as the proportion of correct rejections/true negatives by all items answered incorrectly (Feuerman and Miller 2008). Both scores range from 0 (low sensitivity/specificity) to 1 (high sensitivity/specificity). As the two scores are based on different answer categories (correct and incorrect item answers), it is supposed that sensitivity and specificity either constitute two separate monitoring processes, or that they each measure unique variance in a single monitoring process (Schraw et al. 2013).[4]

$$\text{Sensitivity} = \frac{\sum a}{\sum (a + c)} \tag{2}$$

$$\text{Specificity} = \frac{\sum d}{\sum (b + d)} \tag{3}$$

To test our hypotheses, we conducted hierarchical regression analyses for the metacognitive judgment as well as for each of the accuracy scores.[5] In the first step, we regressed the judgments on gender, prior performance (as measured by high school grade), and test performance. In the second step, we added motivational and personality variables as predictors. The same analyses were performed with the judgment accuracy scores (bias, sensitivity, and specificity), except that we excluded test performance as a predictor variable because the accuracy scores are statistically dependent on performance (Händel and Fritzsche 2016). To test whether regression coefficients for the local and global judgment (or the local and global bias score) differ from each other, we analyzed the interaction term of a general linear mixed model with repeated measures.

Results

Descriptive statistics Students reported a final high school grade of 3.62 ($SD = 0.50$) and a final grade in mathematics of 2.92 ($SD = 1.10$). Descriptive statistics and bivariate correlations for judgments and accuracy scores, performance, motivational, and personality variables are provided in Table 3. For the domain of mathematics, the test was of average difficulty (i.e., students correctly answered about half of the items). The mathematics test score significantly correlated with the final high school grade in mathematics ($r = .47$, $p < .001$), indicating

---

[4] As metacognitive accuracy scores each depend on item performance, and as sensitivity and specificity are composite scores, it is challenging to calculate and interpret their reliability. Still, internal consistency for the judgments, the local bias score, and the signal detection categories based on the 2 × 2 table seem satisfying ($.60 \leq \alpha \leq .89$).
[5] Statistical assumptions for the regression analyses were met (data were not skewed and did not show kurtosis).

Table 3 Study 1: Descriptive statistics and correlations of the variables

Variable	Min	Max	M	SD	1	2	3	4	5	6	7	8	9	10
Mathematics test														
Local judgments														
Judgment	0.00	1.00	0.60	0.25	.07	.10	.48***	.04	.20**	-.04	-.09	.03	-.10	-.02
Bias	-0.33	0.72	0.15	0.20	.14*	-.00	.09	.05	.08	.10	-.04	.03	-.10	-.03
Sensitivity	0.00	1.00	0.67	0.29	.03	.03	.43***	.04	.16*	-.08	-.11	-.04	-.08	.01
Specificity	0.00	1.00	0.41	0.27	-.09	-.06	-.42***	-.01	-.15*	-.06	.09	-.04	.08	.03
Global judgment														
Judgment	-0.06	1.00	0.42	0.27	.07	.01	.57***	.05	.20***	-.13*	-.12	.04	-.10	-.03
Bias	-0.56	0.89	0.02	0.22	.13*	-.10	.25***	.07	.09	-.03	-.08	.04	-.10	-.04
Performance	0.11	1.00	0.46	0.20	-.06	.13*	.49***	-.00	.17**	-.14*	-.07	.01	-.03	.01
Psychology exam														
Global judgment														
Judgment	0.12	0.98	0.68	0.15	-.03	.05	.24***	.30***	.18*	.11	-.03	.28***	.05	.25**
Bias	-0.69	0.28	-0.01	0.15	.06	.04	.08	.17	.09	.18*	-.08	.08	-.12	.10
Performance	0.25	0.94	0.69	0.13	-.09	-.01	.15	.17*	.13	-.04	.04	.25**	.18*	.22*
Motivation and personality														
1 Optimism	3.00	6.00	4.91	0.60	—	-.02	.01	.35	.29***	.38***	.08***	.08	-.42***	.10
2 Narcissism	1.00	6.00	3.32	1.00			-.02	.10	-.03	.10	-.12	-.10	.20**	.19***
3 Domain-specific self-concept	1.00	6.00	3.86	1.28				.06	.23***	-.19**	-.18**	.12*	-.11	.03
4 Academic self-concept	2.20	6.00	4.09	0.67					.42***	.26***	-.03	.15***	-.26***	.34***
5 Self-efficacy	1.00	6.00	4.23	0.74						.08	-.13*	.29***	-.32***	.17**
6 Extraversion	1.50	6.00	4.25	0.88							.10	.03	-.12*	.18*
7 Agreeableness	1.25	5.50	3.72	0.88								.26	.01	-.01
8 Conscientiousness	1.75	6.00	4.14	0.83									-.05	.14*
9 Neuroticism	1.00	6.00	3.66	0.96										.00
10 Openness	1.60	6.00	4.54	0.91										—

Correlations between mathematics and psychology performance scores were as follows: Test performance: $r = .07$, global judgment: $r = .20$, $p < .05$, global bias: $r = .24$, $p < .01$. Correlations between local and global judgments for the mathematics test were as follows: judgment: $r = .61$, $p < .001$, bias $r = .66$, $p < .001$

* $p < .05$. ** $p < .01$. *** $p < .001$

convergent evidence. Students' local judgments were overconfident (positive value for local bias). Descriptive statistics furthermore indicate that students detected correct items more frequently than incorrect items (higher sensitivity than specificity). The exam score in psychology indicates that students performed quite well. The exam score significantly correlated with total final high school grade ($r = .40$, $p < .001$), again indicating convergent evidence. The global bias scores for both the mathematics test and the psychology exam were close to zero, indicating no strong tendency to under- or overestimate test performance. Regarding the implemented motivational and personality scales, mean scores ranged between 3.5 and 5.0, indicating positive values in the respective scales (except narcissism with lower values).

Individual differences in local judgments Results of the hierarchical regression analyses for the local judgments in the mathematics test are reported in Table 4. While none of the investigated predictors contributed to explaining variance in the bias score, 20% to 44% of variance in judgment, sensitivity, and specificity could be explained by the predictors. Performance was found to be the strongest predictor for the judgments. Gender significantly predicted only the level of the judgments as well as the discrimination scores. Female students provided lower local judgments, and had lower sensitivity but higher specificity scores. Among the investigated motivational and personality scales, the only consistent predictor for judgments and metacognitive judgment accuracy was domain-specific self-concept. The higher the self-concept, the higher the judgments, the bias, and the sensitivity, but the lower the specificity. That is, with a higher self-concept, students more often thought that they had answered the items correctly, which was actually an overestimation of their performance. In addition, with higher self-concept, students more often detected items answered correctly but less often detected items answered incorrectly.

Individual differences in global judgments Results regarding global judgments and global bias in both tests (mathematics test and psychology exam) are reported in Table 5. For the

Table 4 Study 1: Separate hierarchical regression analyses for local performance judgments and respective accuracy/discrimination scores in the mathematics test

Predictor	Judgment		Bias		Sensitivity		Specificity	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.39***		.01		.10***		.10***	
Gender		−.11*		−.07		−.21**		.13*
High school grade in mathematics		.09		−.06		.22**		−.28***
Performance		.54***						
Step 2	.05*		.05		.10**		.13***	
Optimism		.08		.11		.04		−.08
Narcissism		.05		.00		−.01		−.10
Domain-specific self-concept		.20*		.10		.37***		−.33***
Academic self-concept		−.03		.00		−.01		.10
Self-efficacy		.05		.00		.07		−.10
Extraversion		.07		.09		.01		−.15*
Agreeableness		.01		−.03		.03		.00
Conscientiousness		.01		.05		−.08		−.01
Neuroticism		−.01		−.03		.03		.00
Openness		−.05		−.09		.00		.07
Total R^2	.44***		.05		.20***		.23***	

* $p < .05$. ** $p < .01$. *** $p < .001$

Table 5 Study 1: Separate regression analyses for global judgments regarding the mathematics test as well as regarding psychology exam

Predictor	Mathematics				Psychology			
	Global judgment		Bias global		Global judgment		Bias global	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.43***		.04**		.24***		.02	
Gender		−.20***		−.18**		.07		.02
High school grade in mathematics		.18***		.06				
Overall high school grade						−.01		−.15
Performance		.47***				.47***		
Step 2	.09***		.09*		.12***		.13*	
Optimism		.14*		.17*		−.02		−.11
Narcissism		−.07		−.14*		.03		.04
Domain-specific self-concept		.31***		.24**		–		–
Academic self-concept		.02		.06		.24*		.19*
Self-efficacy		.02		−.04		−.13		−.13
Extraversion		.00		.00		.13		.16
Agreeableness		.02		−.02		−.12		−.13
Conscientiousness		.04		.08		.27**		.22*
Neuroticism		.07		.08		−.13		−.10
Openness		−.02		−.04		.14		.17
Total R^2	.52***		.13*		.36***		.15*	

Results regarding the psychology exam refer to the subsample for $N = 128$ students. Results regarding the mathematics test refer to the whole sample but remain stable with the subsample as well

$p < .05.$ ** $p < .01.$ *** $p < .001$

global judgments, again, test performance was the strongest predictor, followed by (domain-specific) self-concept. Gender and high school grade as an indicator for prior performance were significant predictors for the judgment and the bias score, but only in the domain of mathematics. Female students provided lower global judgments and underestimated their performance in the mathematics test. Students with better high school grades provided higher judgments. Regarding the mathematics test, optimism additionally predicted the global judgment and overconfidence (and narcissism was negatively related to overconfidence). In contrast, for the exam in psychology, conscientiousness positively predicted the global judgment/global judgment bias.

Individual differences: A comparison between local and global judgments For the mathematics test, judgments were assessed on the local as well as on the global level by the same participants. This allowed comparing beta-weights between local and global judgments. For both judgment and bias, domain-specific self-concept more strongly predicted the global judgments than the local judgments (judgment: $F(1, 243) = 11.04$, $p < .001$, $\eta^2 = .043$; bias: $F(1, 243) = 10.78$, $p < .001$, $\eta^2 = .042$). Significant differences were also found for high school grade (judgment: $F(1, 232) = 6.14$, $p = .014$, $\eta^2 = .026$; bias: $F(1, 232) = 6.93$, $p = .009$, $\eta^2 = .029$), and gender (judgment: $F(1, 242) = 8.81$, $p = .003$, $\eta^2 = .035$; bias: $F(1, 242) = 7.83$, $p = .006$, $\eta^2 = .031$). That is, high school grade (as an indicator for prior performance) and gender more strongly predicted the global judgments and the according bias score. No significantly different relationships for test performance or optimism as further significant predictors on local or global judgments or bias values were found.

Discussion

Study 1 simultaneously investigated the relationship of several performance, motivational, and personality variables with local and global judgments in two domains. Several metacognitive accuracy scores were calculated and regressed on potential determinants derived from the literature. The investigation of the level of the judgment as well as accuracy and discrimination scores provided new insights regarding the question of which factors are relevant for different types and scores of metacognitive judgments. The study found similarities and differences in possible predictors for local and global judgments and respective discrimination scores.

Overall, only selected variables predicted judgments and their accuracy. Test performance was positively related with local and global judgments, which is in line with our first hypothesis as well as with previous research (e.g., Händel and Fritzsche 2016; Kruger and Dunning 1999). Controlling for performance, self-concept was furthermore related to higher judgments, revealing cognitive as well as motivational effects of self-concept. In addition, self-concept was related to higher sensitivity and lower specificity in the mathematics test as well as with more overconfidence in both domains. These results are, however, limited because we assessed domain-specific self-concept only in the domain of mathematics but not in the domain of psychology. Therefore, domain-specific self-concept was collected in Study 2 for the domain of psychology as well.

Besides test performance and self-concept, some further variables were identified as significant predictors, although their effects were not found consistently across the various judgment scores. As expected, motivational and personality variables were more relevant for global than local judgments (H2). Significant differences between local and global judgments were found for domain-specific self-concept, prior performance, and gender. Domain-specific self-concept was more strongly related to global than local judgments and bias. Hence, when students judged their performance regarding a whole test, they seemed to rely more on general beliefs about their domain-specific abilities as information-based cues. In contrast, when judging their performance regarding concrete items, they might rely more on task-specific cues (Koriat 1997; Koriat et al. 2008; Thiede et al. 2010). Karst et al. (2018) who compared global and aggregated local judgments of teachers about students' knowledge highlighted, "the whole is not the sum of its parts" (p. 194). We agree with them and think that differences between local and global judgments might be explained because these judgments are affected differently by possible predictors. In addition, the relatively large number of test items (18 and 32 items) might have implications for the anchor point that students use to generate their judgment. An assumption here could be that with an increasing number of items, students might anchor their judgment on more general cues.

Interestingly, except a small negative effect of extraversion, none of the Big Five factors significantly predicted the judgments or judgment accuracy scores in the mathematics test. Conscientiousness significantly predicted only the global judgment in the final psychology exam as a high stakes condition. When students prepared for a real exam, students' conscientiousness led to a higher judgment of test performance. However, with higher conscientiousness, students were more overconfident. This result seems counterintuitive and was found only in one other study (Schaefer et al. 2004). In general, the lack of effects of the Big Five factors on local as well as global judgments are in contrast to previous research that found at least some (small) effects for some of the Big Five factors (Buratti et al. 2013; Schaefer et al. 2004). In contrast to previous research that mostly implemented general knowledge questions, we used tasks in two specific knowledge domains. This probably allowed students to anchor

their judgments on preconceived notions of their performance in this domain other than personality trait factors.

In sum, the study identified self-concept, which might promote the use of information-based cues, and test performance, which is related to experience-based cues, as relevant predictors for judgments after test taking. This was consistently found for local and global judgments. However, the study did not reveal whether this result transfers to judgments made *before* test taking. Here, students may base their judgment on preconceived beliefs in this domain (e.g., prior achievement, self-concept) but have not had experiences with the test yet. Therefore, Study 2 examined whether the relationships of test performance and self-concept with metacognitive judgments transfers to predictions.

Study 2

Study 2 aimed to replicate the results of Study 1 regarding the psychology exam and extend results to predictions. Based on the findings of the first study, we focused on academic self-concept and we complemented our assessment with domain-specific self-concept regarding the psychology exam.

Design of the study

One week before the course exams took place, students were asked to predict their raw performance score in the course exam (global prediction). In addition, students filled in a questionnaire regarding high school examination grade, gender, academic self-concept, and domain-specific self-concept. In the last week of the term, students participated in a final exam, which was followed by a global judgment (global postdiction).

Sample

The study encompassed 138 undergraduate students from an educational psychology course (73.9% female).[6] Most students were enrolled in their first ($n = 91\%$) study year.

Instruments

High school grade Students' high school grade (possible range from 1 to 6) was assessed via a self-report.

Performance score The final exam of a psychology lecture (same content as in Study 1) served as a performance measure.

Judgments Global prediction and postdiction were each assessed via one item. In the week before the exam, students were instructed to write down the number of items they expected to answer correctly. Directly after finishing the exam, students indicated the number of items they thought they had answered correctly.

---

[6] Data of 17 further students who did not provide a global postdiction were not incorporated in the data.

Self-concept Domain-specific self-concept was assessed via a 5-item scale by Köller et al. (2000), adapted to the domain of psychology (sample item: "Psychology is not for me", $\alpha = .84$). Academic self-concept was assessed by the same scale as in Study 1 by Dickhäuser et al. (2002) with $\alpha = .80$.

Data analyses

For pre- and postdictions, bias as the signed difference of judgment and performance was calculated. After checking for assumptions of normality, we conducted separate hierarchical regression analyses for both judgments (global pre- and postdiction) and the respective two bias scores. In the regression analyses regarding the predicted and postdicted judgment, we included gender, high-school grade as an indicator for prior performance, and test performance in the first step. Domain-specific and academic self-concept were implemented in the second step. Similarly, analyzing potential predictors for prediction bias and postdiction bias, the same variables were implemented in the regression analyses (excluding performance as a predictor because of its statistical dependence on the bias score). Differences of regression coefficients between pre- and postdictions were investigated via the interaction term of a general linear mixed model with repeated measures.

Results

Descriptive values and correlations are provided in Table 6. Overall, students provided higher predictions than postdictions. The level of academic self-concept was similar to that reported in Study 1.

The hierarchical regression analysis to explain variance in the global postdiction replicated the prior results: Test performance was again the strongest predictor, followed by domain-specific self-concept (see Table 7). No significant relations with gender or high school grade were found. In contrast, for the prediction, only the motivational variables in the second step of the regression analysis significantly explained variance in the predicted judgment, with stronger effects for domain-specific than academic self-concept. Regarding the bias scores, significant variance could be only explained for the postdiction bias but nor for the prediction

Table 6 Study 2: Descriptive statistics and correlations of the variables

Variable	Min	Max	M	SD	2	3	4	5	6	7	8
Judgments											
1 Prediction	0.13	0.97	0.70	0.14	.69**	.47**	.42**	.10	.12	.28**	.38**
2 Prediction bias	−0.72	0.41	0.00	0.18		.02	.52**	−.17*	−.63**	.09	.10
3 Postdiction	0.06	0.97	0.63	0.18			.72**	.21	.48**	.26**	.42**
4 Postdiction bias	−0.56	0.25	−0.07	0.16				−.04	−.27**	.15	.24**
Performance											
5 Prior performance	2.40	5.00	3.62	0.54					.34***	.20*	.24**
6 Exam performance	0.28	0.94	0.70	0.13						.18*	.28**
Self-concept											
7 Academic self-concept	1.80	6.00	3.97	0.56							.46**
8 Domain-specific self-concept	2.25	6.00	4.35	0.88							—

$* p < .05. ** p < .01. *** p < .001$

Table 7 Study 2: Separate hierarchical regression analyses for global pre- and postdictions and according bias scores

Predictor	Prediction		Prediction bias		Postdiction		Postdiction bias	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.04		.03		.24***		.02	
Gender		.13		.01		−.10		−.15
High school grade		.05		−.17*		.07		.01
Performance		.08				.47***		
Step 2	.14***		.03		.09***		.07**	
Academic self-concept		.14*		.08		.06		.06
Domain-specific self-concept		.31***		.11		.28***		.25**
Total R^2	.18***		.05		.33***		.09**	

* $p < .05$. ** $p < .01$. *** $p < .001$

bias. Again, domain-specific self-concept was the strongest predictor: with higher self-concept, students were more overconfident in their postdiction.

Although academic and domain-specific self-concept explained more variance for prediction than for postdiction, none of the direct comparisons of beta-weights were significantly different ($.12 \le p \le .55$). Vice versa, however, test performance was of higher relevance for postdictions than for predictions ($p < .001$). That is, Hypothesis 3 could be partly confirmed.

Discussion

The study replicated and extended the results of Study 1. Again, global postdictions were predicted by test performance and domain-specific self-concept. Similar to Study 1, students with higher self-concepts provided higher judgments.

In addition, the study indicated that this pattern of results was not one-to-one transferrable to predictions. While the global postdiction in an exam setting was most strongly predicted by test performance, followed by domain-specific self-concept, the global prediction was only predicted by academic and domain-specific self-concept. When students had not yet processed the task, their judgment relied mostly on domain-specific self-concept. Test performance itself had no predictive power for predictions. That is, students with different scores on the exam did not significantly differ in their prediction one week before taking the exam. This result contrasts previous research on predictions for real exams in the same domain (Miller and Geraci 2011b; Serra and DeMarree 2016). One reason for the differences between pre- and postdictions might relate to desired performance (Saenz et al. 2017; Serra and DeMarree 2016), probably being more relevant for predictions than postdictions (Händel and Bukowski 2019). However, in our study, there were no available data for students' desired grades.

Results regarding predictions might be limited because they had already been assessed one week before the exam. Hence, an assessment closer to the exam situation may lead to different results because, then, students have more knowledge about how well they have actually prepared for the exam. However, when assessing predictions shortly before exam taking, students might be nervous or anxious, and therefore may not make serious judgments. Furthermore, Study 2 focused only on self-concept as a predictor and, hence, it does not provide any evidence on the relationship of other learner characteristics with predictions. For replication purposes, it would be worthwhile to study the other personality variables as well. A

further limitation is that when regressing predictions on performance, the judgment is regressed on a variable (performance) that is only assessed afterwards.

General discussion

This paper reported two studies that simultaneously investigated possible predictors for several metacognitive judgments and judgment accuracy. The studies were concerned with the relationship of performance, motivation, and personality with different types of judgments and judgment accuracy scores. A unique feature of both studies is that they controlled for the relationship with performance when predicting judgments by motivational and personality variables. Furthermore, to obtain comprehensive insights into the bases of monitoring judgments and their accuracy, we considered not only judgments but also different measures of judgment accuracy (relative and absolute). Thereby, we provide important empirical evidence on the nature of diagnostic cues at the individual level (Koriat et al. 2008). Still, it needs to be noted that due to dependencies between the judgments and the test score, the test score was excluded from the analyses, and only high school grade was used as an indicator for prior performance.

Consistently over both domains and settings, students' global judgments were significantly predicted by test performance. That is, with higher performance, students provided higher global judgments. Over and above the well-known-effect of performance (Kruger and Dunning 1999), self-concept significantly contributed to judgments. While previous research has pointed to relations between self-concept and calibration (Lösch et al. 2017), the current work is the first to study the relationship of self-concept with different metacognitive judgments and measures. Across studies, we showed that with higher self-concept, students provided higher and more overestimated judgments, which is in line with model assumptions by Kröner and Biermann (2007) and others. Moreover, the results of Study 2 indicate that making predictions is differentially affected by performance and motivational variables than making postdictions. When students provide predictions, they have not yet had any experiences with the exam (Bol and Hacker 2012). Hence, their domain-specific self-concept seemed to guide their metacognitive judgments. In contrast, after exam taking, students can shift their evaluation basis to the exam experience, and thereby additionally use experience-based cues (Koriat et al. 2008).

Interestingly, higher self-concept led to too high judgments. Still, we do not know whether the reported self-concepts were realistic or rather optimistic (Praetorius et al. 2016). Overoptimistic self-concepts or self-enhancement tendencies might be responsible for the effect of self-concept on overconfidence (Alicke and Sedikides 2009; Guay et al. 2010; Jiang and Kleitman 2015; Krueger 2016; Marsh and Craven 2006; Marsh and Martin 2011). That is, students who are biased in their self-views and who have the tendency to maintain an unrealistically positive self-view might provide overestimated judgments.

In Study 1, results seemed to differ between the laboratory setting with the mathematics test and the exam setting in a psychology course. In contrast to previous research that either investigated global judgments in exam settings (e.g. Serra and DeMarree 2016) or local judgments in laboratory studies (e.g. Schaefer et al. 2004), our study provided a broader set of settings and judgments. However, as setting (laboratory versus real exam) and domain (mathematics versus psychology) were confounded, we cannot conclude whether the domain or the setting were responsible for different results of possible predictors on global judgments

and bias. Hence, the setting might influence judgments and their accuracy. Laboratory situations have only low personal relevance for the students while exam situations represent high stakes conditions, for which students usually dedicatedly prepare themselves and study the contents (Cogliano et al. 2019). In contrast, regarding tests used in empirical studies, students are usually not even briefed about the topics beforehand. This might have implications for students' engagement in providing accurate judgments and for the cues considered for their judgments (e.g., students might provide more conscientious judgments when the content or the outcome of a test is personally relevant to them). For example, some studies indicated that judgment accuracy differs when students are awarded or graded to provide accurate judgments (Dutke et al. 2010; Gutierrez and Schraw 2015; Miller and Geraci 2011a).

Regarding our study, we suppose that some differences might rather be explained via the domain—like the relationship with gender in mathematics as a stereotypically male-dominated domain (Morony et al. 2013), while others seem be related to the setting. For example, conscientiousness significantly predicted the global judgment in a real exam, for which students should have studied and prepared. In contrast, optimism predicted global judgments in the laboratory setting, in which students had not prepared for the test. To unravel the results, future research is needed that investigates judgments separately in different domains but the same setting and judgments within the same domain but different settings. Gender and high school grades were significant predictors only in the mathematics test but not in the psychology exam. As discussed in Study 1, gender as a significant predictor in mathematics might be explained via gender-specific self-concept in mathematics (Ackerman et al. 2002; Marsh and Yeung 1998; Morony et al. 2013). Prior performance (operationalized via high school grade) was a significant predictor only for the global judgment regarding the mathematics test but not regarding the psychology exam. This difference might be traced back to the domain-specificity of the assessed high school grade (high school grade in mathematics versus overall high school grade). Still, using prior performance as a predictor seems promising from a methodological point of view because it overcomes the pitfall of being statistically dependent on absolute and relative scores, which is the case when using performance of the criterion test as a predictor variable (Kelemen et al. 2007).

Limitations and prospects for future research

Our studies are limited in several regards. First, the complete set of predictors was only implemented in Study 1. For predictions, which were studied in Study 2, only a selection of variables was collected. As predictions in particular might be related with personality variables, future research should consider possible motivational and personality variables when investigating individual differences in predictions. In addition, although we aimed to encompass relevant predictors investigated in earlier research, our approach cannot be regarded as all encompassing. For example, in both studies, we focused on one testing occasion and, hence, did not consider previous judgments as an influencing factor for performance judgments. Moreover, due to its relationship with performance, anxiety might also relate to judgments (Jiang and Kleitman 2015; Morony et al. 2013). A further interesting approach would be to assess self-efficacy specific to content requirements. For example, the PISA study measured mathematics self-efficacy via items assessing "understanding graphs" or "solving specific types of equations" (OECD 2013). Second, comparison of results regarding self-concepts between domains in Study 1 is limited as it was assessed domain-specifically only in mathematics. Still, Study 2 confirmed that for each dependent variable, domain-specific self-concept was a stronger predictor than academic self-concept. Finally, some of the predictors used in our regression analyses, especially performance and self-concept were

significantly interrelated (see also Marsh 1990), which might have influenced the results regarding the judgments, where performance was used as a predictor variable in the first step of the regression. While our study aimed to assess judgments related to a personally relevant performance criterion, for future research, it seems worthwhile to use external performance criteria—assessed via a separate test as in the study by Kelemen et al. (2007). Such research would also overcome the limitation of using convenience samples as in our studies, which were implemented in course settings. In addition, an interesting prospect for future research would be to study the interrelation between judgments and performance, that is, whether performance scores in one criterion test influence metacognitive judgments regarding another criterion test, and vice versa (Foster et al. 2017).

Conclusion

The current two studies covered a broad set of possible predictors of metacognitive judgments. Test performance and domain-specific self-concept were identified as the most important predictors. However, differences between varying types of judgments suggest that research on individual differences needs to distinguish between judgments of different grain size and the different points in time they are provided at. In particular, self-concept was a stronger predictor for global judgments than for local judgments. This suggests that self-concept as a possible predictor might have been underestimated in most previous studies because they focused on local judgments when investigating individual differences. Similarly, future research clearly needs to distinguish whether students judge their performance before or after test completion, as domain-specific and academic self-concept explained more variance for predictions. Using our results as a starting point, a promising prospect for future research seems to be to conceptually and empirically combine research on self-concept with research on metacognitive judgments (Lösch et al. 2017).

Compliance with ethical standards

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

References

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences, 33*, 587–605. https://doi.org/10.1016/S0191-8869(01)00174-X.

Agler, L.-M. L., Noguchi, K., & Alfsen, L. K. (2019). Personality traits as predictors of reading comprehension and metacomprehension accuracy. *Current Psychology*, 1–10. https://doi.org/10.1007/s12144-019-00439-y.

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*, 1–48. https://doi.org/10.1080/10463280802613866.

Ariel, R., Lembeck, N. A., Moffat, S., & Hertzog, C. (2018). Are there sex differences in confidence and metacognitive monitoring accuracy for everyday, academic, and psychometrically measured spatial ability? *Intelligence, 70*, 42–51. https://doi.org/10.1016/j.intell.2018.08.001.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*, 122–147. https://doi.org/10.1037/0003-066x.37.2.122.

Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2012). Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen: Allgemeine Selbstwirksamkeit Kurzskala (ASKU) [assessing subjective competency beliefs: A short version to assess general self-efficacy]. In *GESIS-Working Papers* (Vol. 17). Mannheim: GESIS.

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education, 69*, 133–151. https://doi.org/10.1080/00220970109600653.

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

de Bruin, A. B. H., Kok, E. M., Lobbestael, J., & de Grip, A. (2017). The impact of an online tool for monitoring and regulating learning at university: Overconfidence, learning strategy, and personality. *Metacognition and Learning, 12*, 21–43. https://doi.org/10.1007/s11409-016-9159-5.

Buratti, S., Allwood, C. M., & Kleitman, S. (2013). First- and second-order metacognitive judgments of semantic memory reports: The influence of personality traits and cognitive styles. *Metacognition and Learning, 8*, 79–102. https://doi.org/10.1007/s11409-013-9096-5.

Burns, K. M., Burns, N. R., & Ward, L. (2016). Confidence - more a personality or ability trait? It depends on how it is measured: A comparison of young and older adults. *Frontiers in Psychology, 7*, 518. https://doi.org/10.3389/fpsyg.2016.00518.

Campbell, W. K., Goodie, A. S., & Foster, J. D. (2004). Narcissism, confidence, and risk attitude. *Journal of Behavioral Decision Making, 17*, 297–311. https://doi.org/10.1002/bdm.475.

Carver, C. S., Scheier, M. F., & Segerstrom, S. C. (2010). Optimism. *Clinical Psychology Review, 30*, 879–889. https://doi.org/10.1016/j.cpr.2010.01.006.

Cogliano, M. C., Kardash, C. A. M., & Bernacki, M. L. (2019). The effects of retrieval practice and prior topic knowledge on test performance and confidence judgments. *Contemporary Educational Psychology, 56*, 117–129. https://doi.org/10.1016/j.cedpsych.2018.12.001.

Dahl, M., Allwood, C. M., Rennemark, M., & Hagberg, B. (2010). The relation between personality and the realism in confidence judgements in older adults. *European Journal of Ageing, 7*, 283–291. https://doi.org/10.1007/s10433-010-0164-2.

Dickhäuser, O., Schöne, C., Spinath, B., & Stiensmeier-Pelster, J. (2002). Die Skalen zum akademischen Selbstkonzept [The academic self concept scales]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 23*, 393–405. https://doi.org/10.1024//0170-1789.23.4.393.

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4–14. https://doi.org/10.1016/j.learninstruc.2012.06.001.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension. *Current Directions in Psychological Science, 16*, 228–232. https://doi.org/10.1111/j.1467-8721.2007.00509.x.

Dutke, S., Barenberg, J., & Leopold, C. (2010). Learning from text: Knowing the test format enhanced metacognitive monitoring. *Metacognition and Learning, 5*, 195–206. https://doi.org/10.1007/s11409-010-9057-1.

Efklides, A., & Tsiora, A. (2002). Metacognitive experiences, self-concept, and self-regulation. *Psychologia, 45*, 222–236. https://doi.org/10.2117/psysoc.2002.222.

Farwell, L., & Wohlwend-Lloyd, R. (1998). Narcissistic processes: Optimistic expectations, favorable self-evaluations, and self-enhancing attributions. *Journal of Personality, 66*, 65–83. https://doi.org/10.1111/1467-6494.t01-2-00003.

Feuerman, M., & Miller, A. R. (2008). Relationships between statistical measures of agreement: Sensitivity, specificity and kappa. *Journal of Evaluation in Clinical Practice, 14*, 930–933. https://doi.org/10.1111/j.1365-2753.2008.00984.x.

Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: The role of memory for past exam performance in student predictions. *Metacognition and Learning, 12*, 1–19. https://doi.org/10.1007/s11409-016-9158-6.

Fritzsche, E. S., Kröner, S., Dresel, M., Kopp, B., & Martschinke, S. (2012). Confidence scores as measures of metacognitive monitoring in primary students? (limited) validity in predicting academic achievement and the mediating role of self-concept. *Journal for Educational Research Online, 4*, 120–142.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528. https://doi.org/10.1037/0033-295X.98.4.506.

Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences, 20*, 644–653. https://doi.org/10.1016/j.lindif.2010.08.001.

Gutierrez, A. P., & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration. *The Journal of Experimental Education, 83*, 386–404. https://doi.org/10.1080/00220973.2014.907230.

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*, 160–170. https://doi.org/10.1037//0022-0663.92.1.160.

Händel, M., & Bukowski, A.-K. (2019). The gap between desired and expected performance as predictor for judgment confidence. *Journal of Applied Research in Memory and Cognition, 8*, 347–354. https://doi.org/10.1016/j.jarmac.2019.05.005.

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

Händel, M., & Fritzsche, E. S. (2015). Students' confidence in their performance judgements: A comparison of different response scales. *Educational Psychology, 35*, 377–395. https://doi.org/10.1080/01443410.2014.895295.

Händel, M., & Fritzsche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory & Cognition, 44*, 229–241. https://doi.org/10.3758/s13421-015-0552-0.

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*. https://doi.org/10.1016/j.learninstruc.2019.101245.

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences, 30*, 11–21. https://doi.org/10.1016/j.lindif.2013.12.003.

Jiang, Y., & Kleitman, S. (2015). Metacognition and motivation: Links between confidence, self-protection and self-enhancement. *Learning and Individual Differences, 37*, 222–230. https://doi.org/10.1016/j.lindif.2014.11.025.

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206–219. https://doi.org/10.1037/0022-3514.66.1.206.

Jonsson, A.-C., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences, 34*, 559–574. https://doi.org/10.1016/s0191-8869(02)00028-4.

Karst, K., Dotzel, S., & Dickhäuser, O. (2018). Comparing global judgments and specific judgments of teachers about students' knowledge: Is the whole the sum of its parts? *Teaching and Teacher Education, 76*, 194–203. https://doi.org/10.1016/j.tate.2018.01.013.

Kasperski, R., & Katzir, T. (2013). Are confidence ratings test- or trait-driven? Individual differences among high, average, and low comprehenders in fourth grade. *Reading Psychology, 34*, 59–84. https://doi.org/10.1080/02702711.2011.580042.

Kelemen, W. L., Winningham, R. G., & Weaver, C. A. (2007). Repeated testing sessions and scholastic aptitude in college students' metacognitive accuracy. *European Journal of Cognitive Psychology, 19*, 689–717. https://doi.org/10.1080/09541440701326170.

Köller, O., Schnabel, K. U., & Baumert, J. (2000). Der Einfluß der Leistungsstärke von Schulen auf das fachspezifische Selbstkonzept der Begabung und das Interesse [The influence of performance of schools on domain-specific self-concept and interest]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 32*, 70–80. https://doi.org/10.1026//0049-8637.32.2.70.

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

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 117–135). New York: Psychology Press.

Kornell, N., & Hausman, H. (2017). Performance bias: Why judgments of learning are not affected by learning. *Memory & Cognition, 45*, 1270–1280. https://doi.org/10.3758/s13421-017-0740-1.

Kröner, S., & Biermann, A. (2007). The relationship between confidence and self-concept — Towards a model of response confidence. *Intelligence, 35*, 580–590. https://doi.org/10.1016/j.intell.2006.09.009.

Krueger, J. (2016). Enhancement bias in descriptions of self and others. *Personality and Social Psychology Bulletin, 24*, 505–516. https://doi.org/10.1177/0146167298245006.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology, 82*, 180–188. https://doi.org/10.1037//0022-3514.82.2.180.

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*, 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121.

Küfner, A. C. P., Dufner, M., & Back, M. D. (2015). Das Dreckige Dutzend und die Niederträchtigen Neun [Dirty Dozen and Naugthy Nine]. *Diagnostica, 61*, 76–91. https://doi.org/10.1026/0012-1924/a000124.

Lienert, G. A., & Hofer, M. (1972). *MTAS. Mathematiktest für Abiturienten und Studienanfänger [Mathematics test for high-school graduates and freshmen]*. Göttingen: Hogrefe.

Lösch, T., Kelava, A., Nagengast, B., Trautwein, U., & Lüdtke, O. (2017). Perspective matters: The internal/external frame of reference model for self- and peer ratings of achievement. *Learning and Instruction, 52*, 80–89. https://doi.org/10.1016/j.learninstruc.2017.05.001.

Luszczynska, A., Scholz, U., & Schwarzer, R. (2005). The general self-efficacy scale: Multicultural validation studies. *The Journal of Psychology, 139*, 439–457. https://doi.org/10.3200/JRLP.139.5.439-457.

Macenczak, L. A., Campbell, S., Henley, A. B., & Campbell, W. K. (2016). Direct and interactive effects of narcissism and power on overconfidence. *Personality and Individual Differences, 91*, 113–122. https://doi.org/10.1016/j.paid.2015.11.053.

Marsh, H. W. (1986). Self-serving effect (bias?) in academic attributions: Its relation to academic achievement and self-concept. *Journal of Educational Psychology, 78*, 190–200. https://doi.org/10.1037/0022-0663.78.3.190.

Marsh, H. W. (1990). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2*, 77–172. https://doi.org/10.1007/BF01322177.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*, 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x.

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 81*, 59–77. https://doi.org/10.1348/000709910X503501.

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20*, 107–123. https://doi.org/10.1207/s15326985ep2003_1.

Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and english constructs. *American Educational Research Journal, 35*, 705–738. https://doi.org/10.3102/00028312035004705.

McCrae, R., & Costa, P. J. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81–90. https://doi.org/10.1037/0022-3514.52.1.81.

McCrae, R., & Costa, P. J. (1999). A five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 139–153). New York: The Guilford Press.

Miller, T. M., & Geraci, L. (2011a). 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.

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

Morony, S., Kleitman, S., Lee, Y. P., & Stankov, L. (2013). Predicting achievement: Confidence vs self-efficacy, anxiety, and self-concept in Confucian and European countries. *International Journal of Educational Research, 58*, 79–96. https://doi.org/10.1016/j.ijer.2012.11.002.

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. *The Journal of Experimental Education, 74*, 7–28.

OECD. (2013). *PISA 2012 results: Ready to learn (volume III): Students' engagement, drive and self-beliefs*. Paris: PISA, OECD Publishing.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*, 193–203. https://doi.org/10.1037/0022-0663.86.2.193.

Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles, 48*, 265–276. https://doi.org/10.1023/A:1022877405718.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *The Journal of General Psychology, 129*, 257–299. https://doi.org/10.1080/00221300209602099.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36*, 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6.

Perfect, T. J., & Schwartz, B. L. (Eds.). (2002). *Applied metacognition*. Cambridge: Univ. Press.

Pierce, B. H., & Smith, S. M. (2001). The postdiction superiority effect in metacomprehension of text. *Memory & Cognition, 29*, 62–67. https://doi.org/10.3758/BF03195741.

Praetorius, A.-K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? [do students with optimistic self-evaluations have a slight advantage?]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 48*(1), 14–26. https://doi.org/10.1026/0049-8637/a000140.

Rammstedt, B., & John, O. P. (2005). Kurzversion des big five inventory (BFI-K) [short version of the big five inventory (BFI-K)]. *Diagnostica, 51*, 195–206. https://doi.org/10.1026/0012-1924.51.4.195.

Retelsdorf, J., Schwartz, K., & Asbrock, F. (2015). "Michael can't read!" teachers' gender stereotypes and boys' reading self-concept. *Journal of Educational Psychology, 107*, 186–194. https://doi.org/10.1037/a0037107.

Roelle, J., Schmidt, E. M., Buchau, A., & Berthold, K. (2017). Effects of informing learners about the dangers of making overconfident judgments of learning. *Journal of Educational Psychology, 109*, 99–117. https://doi.org/10.1037/edu0000132.

Rutherford, T. (2017). The measurement of calibration in real contexts. *Learning and Instruction, 47*, 33–42. https://doi.org/10.1016/j.learninstruc.2016.10.006.

Saenz, G. D., Geraci, L., Miller, T. M., & Tirso, R. (2017). Metacognition in the classroom: The association between students' exam predictions and their desired grades. *Consciousness and Cognition, 51*, 125–139. https://doi.org/10.1016/j.concog.2017.03.002.

Satow, L., & Schwarzer, R. (1999). Optimismus (OPTIM) [Optimism]. In R. Schwarzer & M. Jerusalem (Eds.), *Skalen zur Erfassung von Lehrer- und Schülermerkmalen* (pp. 40–41). Berlin: Freie Universität.

Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the big five. *Journal of Research in Personality, 38*, 473–480. https://doi.org/10.1016/j.jrp.2003.09.010.

Schmidt, G. (2013). Hochschulen in Bayern [universities in Bavaria]. *Bayern in Zahlen. Fachzeitschrift für Statistik, 7*, 395–402.

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

Schraw, G., Kuch, F., & Gutierrez, A. P. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction, 24*, 48–57. https://doi.org/10.1016/j.learninstruc.2012.08.007.

Serra, M. J., & DeMarree, K. G. (2016). Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. *Memory & Cognition, 44*, 1127–1137. https://doi.org/10.3758/s13421-016-0624-9.

Sharpe, J. P., Martin, N. R., & Roth, K. A. (2011). Optimism and the big five factors of personality: Beyond neuroticism and extraversion. *Personality and Individual Differences, 51*, 946–951. https://doi.org/10.1016/j.paid.2011.07.033.

Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences, 22*, 747–758. https://doi.org/10.1016/j.lindif.2012.05.013.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes, 47*, 331–362. https://doi.org/10.1080/01638530902959927.

Williams, D. (1992). Dispositional optimism neuroticism and extraversion. *Personality and Individual Differences, 13*, 475–477. https://doi.org/10.1016/0191-8869(92)90076-2.

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). Hillsdale: Erlbaum.