

Developmental Relationship Between Metacognitive Monitoring and Reading Comprehension

Ashenafi K. Edossa¹, Kathrin Lockl² & Sabine Weinert³

¹ Bamberg Graduate School of Social Science (BAGSS), University of Bamberg, Bamberg, Germany

² Leibniz Institute for Educational Trajectories, University of Bamberg, Bamberg, Germany

³ Department of Developmental Psychology, University of Bamberg, Bamberg, Germany

Correspondence: Ashenafi K. Edossa, University of Bamberg, Feldkirchenstr. 21, Bamberg, 96052, Germany.
E-mail: ashenafieldossa@gmail.com

Received: September 11, 2022 Accepted: November 23, 2022 Online Published: November 30, 2022

doi:10.5539/jedp.v13n1p1

URL: <http://doi.org/10.5539/jedp.v13n1p1>

Abstract

Metacognitive processes are crucial components of self-regulatory ability. The study examined the developmental relationship between metacognitive monitoring and reading comprehension using longitudinal data from the National Educational Panel Study (NEPS) in Germany. We used a cross-lagged panel analysis – in the framework of Structural Equation Modeling – to examine the reciprocal relationship between metacognitive monitoring (assessed as a retrospective judgement of one’s performance) and reading comprehension at three time points: grades 5, 7, and 9. For the overall group of participants, we consistently found longitudinal reciprocal effects between metacognitive monitoring and reading comprehension from grade 5 to 7 and grade 7 to 9. This implies monitoring accuracy leads to higher reading comprehension over time and vice versa. Although the accuracy of students’ judgement was positively and reciprocally related to reading comprehension, a multi-group cross-lagged panel analyses showed that excessive underestimation appeared to have a more hindering effect than excessive overestimation on later reading comprehension. However, moderate underestimation might be less detrimental than excessive overestimation. In addition to shedding light on the scientific debate on the complex developmental interplay between reading comprehension and metacognitive monitoring, theoretical and practical implications of the results are provided.

Keywords: Procedural metacognition, metacognitive monitoring, judgement accuracy, monitoring accuracy, reading comprehension, development, self-regulation

1. Introduction

Metacognitive processes are components of self-regulatory abilities that help individuals to cope with everyday life challenges. Knowing about, monitoring, and controlling one’s cognitive enterprise is important to effectively solve problems in and outside schools (Flavell, 1979; Schraw & Moshman, 1995). Metacognition is “any knowledge or cognitive activity that takes as its object or regulates, any aspect of any cognitive enterprise” (Flavell & Miller, 2002, p. 150). Paris and Byrnes (1989) differentiated between declarative and procedural metacognition. Declarative metacognition refers to conscious and detailed knowledge about factors in tasks and situational variables that affect cognitive performance and the reasons behind that (Paris, 2002; Edossa et al., 2019). The declarative facet of metacognition is mainly “verbalizable, stable, and late-developing” (Schneider, 2015, p. 258). Procedural metacognition, on the other hand, is the application of metacognitive aspects such as monitoring, controlling, and regulating (Fritz et al., 2010; Nelson & Narens, 1994).

Monitoring is one component of procedural metacognitive processes and refers to a live awareness of task performance such as text comprehension (Schraw & Moshman, 1995). Nelson and Narens (1990) distinguished between *prospective* monitoring and *retrospective* monitoring. The three common types of prospective monitoring are namely: ease-of-learning, judgment-of-learning, and feeling-of-knowing (Nelson & Narens, 1994). On the other hand, confidence judgement is one aspect of retrospective monitoring in which individuals reflect on the accuracy of a previous recall response (Nelson & Narens, 1990). The focus of the present study is *retrospective monitoring*, specifically, judgements that follow cognitive performance – commonly known as *postdictions* – and its developmental interrelationship with reading comprehension.

1.1 Metacognitive Monitoring and Academic Achievement

The role of self-regulation such as emotional and behavioral regulation (Edossa et al., 2017) and metacognition (Schneider & Artelt, 2010; Tibken et al., 2021; Edossa et al., 2019) on academic achievement is established. Specifically, the concept of metacognition seems applicable in the explanation of children's production deficiencies in several cognitive tasks (Schneider, 2015). Production deficiency is children's ability to deploy a strategy that is taught but an inability to execute spontaneously (for review see Schneider & Pressley, 1997). Metacognition has been positively linked with academic achievement in general and in the process of reading in particular (Pintrich, 2002; Schneider & Pressley, 1997; Zargar et al., 2020). Metacognition is assumed to play a fundamental role in the reflection of live reading processes, choosing and employing relevant strategies (Artelt & Schneider, 2015; Soto et al., 2020). Students who are capable of regulating their learning process are actively engaged in metacognitive activities before and after reading to construct meaning from a text (Paris et al., 1991; Schneider & Pressley, 1997). Like other facets of metacognition, a positive association has been reported between procedural metacognition and academic achievement (Schneider & Pressley, 1997).

Specifically, the monitoring aspect of procedural metacognition plays a vital role in the learning process by providing information about the current epistemic state to regulate the process of learning, effectively (Nelson & Narens, 1994). Metacognitive judgment in the process of monitoring helps us to observe and reflect on our cognitive process (Flavell, 1979). It has a valuable role in self-regulated learning in terms of the allocation of time and employment of appropriate learning strategies (Thiede & Dunlosky, 1999). The decision to continue a study item widely depends on the feedback from the judgement accuracy of students in the monitoring process (Hacker et al., 2000; Thiede & Dunlosky, 1999). Premature termination of study and unnecessarily prolonged study time can be prevented when students acquire the skill to accurately judge their learning process and performance (Hacker et al., 2000). Accurate predictors focus and exert their energy on the right items because they have an insight into the cognitive demand of the task and their ability to perform the given task (Dunning et al., 2003).

However, the relationship between metacognition and academic achievement is not necessarily unidirectional. There is a line of theoretical assumptions that the developmental relationship between metacognition and academic achievement can be reciprocal (Flavell & Wellman, 1977; Schneider & Pressley, 1997). Dunning et al. (2003) assume the ability required to accurately judge one's performance is similar to the skill to correctly solve a given cognitive task. Therefore, they argue that if people fail to produce correct responses, they are double cursed with the inability to be aware of whether their responses are correct or wrong. For instance, incompetent readers fail to understand what they comprehend from the text (Dunning et al., 2003). This implies the importance of knowledge for metacognitive monitoring process.

Tibken et al. (2021) investigated the effects of metacognitive competence – declarative and procedural metacognition—on the development of school achievement in gifted and non-gifted adolescents, and the result of the study showed that metacognitive competence predicted school achievement (including reading comprehension) for both group of participants. The study linked low metacognitive competence to underachievement among gifted children. Roebers et al. (2014) demonstrated a strong and positive effect of metacognitive monitoring on test performance in a study conducted among 9 and 11 years old children with $\beta = .76$ and $\beta = .52$, respectively. Similarly, Stankov et al. (2012) found the accuracy of judgement to be the best predictor of mathematics and reading achievement. Roderer and Roebers (2010) found that the highest achieving students were most accurate in performance prediction. Overall, the positive role of metacognitive competence on academic achievement seems to be well established (Bouffard et al., 1998; Tibken et al., 2021; Roebers et al., 2014). On the other hand, King and McInerney (2016) examined the longitudinal relationship between metacognitive strategy (changing to improve, monitoring and planning) and academic achievement (mathematics and reading comprehension). They employed self-reported measures to assess metacognitive processes and cross-lagged analysis to investigate the developmental interplay between metacognitive strategy and academic achievement. However, they found a contradictory result to the established literature that metacognitive processes had no statistically significant longitudinal effects on academic achievement. The study revealed a uni-directional influence from academic achievement to metacognitive strategy only.

Cai et al. (2019) studied the developmental interplays among goal, metacognitive strategy, and school achievement longitudinally across three time points, and the authors reported a bi-directional developmental interplay between metacognitive strategy and academic achievement from time points 1 to 2. However, the developmental interplay became uni-directional from time points 2 to 3. Only metacognitive strategy had a statistically significant prediction on academic achievement in the later time points. Taken together, the literature on the developmental association between metacognitive monitoring and academic achievement appears inconsistent. Therefore, the

main aim of the present study is to examine the developmental relationship between metacognitive monitoring and reading comprehension by taking into account socioeconomic status (SES) since past investigations have shown a positive effect of SES on both metacognition and achievement (Thompson & Foster, 2014; Yerdelen-Damar & Peşman, 2013). Meta-analytic studies reported that SES and academic achievement have a moderate to strong relationship (Sirin, 2005; Liu, Peng, & Luo, 2020). Metacognition was reported to play a mediator role in the relationship between SES and academic achievement (e.g., Yerdelen-Damar et al., 2013). SES might also moderate the relationship between metacognition and academic achievement (e.g., Koyuncu et al., 2021). Therefore, SES is controlled to rule out the confounding effects of SES on the result of the study.

1.2 Implications of Under-and Overestimation on Achievement

The role of monitoring accuracy in academic achievement has been discussed from a metacognitive perspective. Low monitoring accuracy might be due to under – or overestimation. Researchers from the perspective of social psychology, personality, and motivation further debate the implication of under-and overestimation of one's performance in a healthy developmental adaptation of children. Overall, people tend to overestimate their competence in several areas such as reasoning, grammatical ability, and social skills (Kruger & Dunning, 1999; Dunning et al., 2003). Generally, high confidence is observed for correct answers across all age groups but young children tend to be more overconfident than older children and adults for answers that turn out to be incorrect (Roderer & Roebbers, 2010). While the metacognitive research perspective emphasizes the importance of judgement accuracy for effective regulation of learning, the other perspectives from social psychology, personality, and motivation studies have mixed assumptions on the adaptive implication of under – and overestimation in and outside school (Bouffard & Narciss, 2011). Drawing on assumptions from social cognitive theory, some argue that overestimation of one's performance is useful for academic competence because it boosts motivation and enhances persistence in challenging contexts (Bouffard et al., 2006; Bouffard & Narciss, 2011). They argue overestimation can be helpful to solve novel and difficult tasks by making students persistent and motivated – which they would not pursue if they were accurate in judging their competence (Bjorklund & Bering, 2002; Shin et al., 2007). As a result, consistent underestimation of performance is assumed to be limiting and detrimental to academic achievement in terms of motivation (Bouffard & Narciss, 2011). Contrary to this assumption, the other line of the literature suggests that overestimation may be an obstacle for students to identify their learning needs and thus incapable of effectively regulating their learning process and putting the required effort into their study (Dunlosky et al., 2005; Butler & Winne, 1995). Therefore, students who overestimate their performance might not be well prepared, and might not ask for help from other persons (Stone & May, 2002) and thus may achieve lower than students who moderately underestimate their achievement. In addition, because low achieving students may not notice their poor performance, they might engage in overestimation (Dunning et al., 2003). As reflected by the theories, empirical evidence is mixed. For instance, Bouffard et al. (2011) positively associated overestimation with higher academic achievement. In contrast, other studies linked underestimation to high academic achievement (Chiu & Klassen, 2010; Gonida & Leondari, 2011).

Studies conducted from the perspective of metacognition focus on the positive role of monitoring accuracy on academic performance and vice versa. Other researchers assume overestimation or self-enhancement of one's performance could have a positive effect on academic achievement via motivation while underestimation might be detrimental. The difference across different research perspectives is not only in theoretical assumptions but also in terminologies and methods used. Studies conducted in light of social learning theory use social comparison or social consensus as a criterion to determine self-evaluation bias (Bouffard & Narciss, 2011). However, this method is criticized because it is dependent on the attitude of other people (Kwan et al., 2008). To deal with this limitation, metacognition researchers use specific assessment criteria such as test scores to determine the accuracy of judgement of one's competence. There are two popular methods to measure the accuracy of judgement of performance about test scores. The first method calculates a deviation score by subtracting the criterion (actual test score) from self-evaluation/estimation. However, this method has been criticized, especially, in studies that focus on the relationship between judgement and academic competence because the deviation score is confounded with competence (Dufner et al., 2015). To overcome this limitation of deviation score, a second method is recommended which is computed by regressing the self-evaluation (estimation of one's performance) on the criterion (actual test score) and taking the residual score to measure the level of judgment accuracy (Dufner et al., 2015; Gonida & Leondari, 2011). Therefore, in the present study, the residual score was mainly used to measure accuracy of judgement to find out its longitudinal relationship with reading comprehension using large-scale data from the German National Educational Panel Study (NEPS).

1.3 Hypothesis

The literature on procedural metacognition, metacognitive monitoring in particular, and its role in academic achievement are mixed. Researchers from the perspective of metacognition emphasized the importance of monitoring accuracy to effectively employ appropriate strategies and allocate proper time and energy for academic success. In addition, academic achievement is assumed to have effects on the development of metacognitive monitoring accuracy despite scant longitudinal empirical investigations. In addition, it is not clear whether the positive role of monitoring accuracy on achievement in general, reading comprehension in particular, holds true for both under – and overestimating students. Therefore, the study intends to fill this gap by testing the following hypothesis using a residual based score of judgement, which is less confounded to achievement.

- Metacognitive monitoring has a positive effect on the development of reading comprehension from grades 5 to 7 and grades 7 to 9.
- In turn, reading comprehension has a positive effect on the development of metacognitive monitoring from grades 5 to 7 and grades 7 to 9.
- In addition, we aimed to test whether the above hypotheses hold true for under-and overestimating students or not.

2. Method

2.1 Participants and Procedures

The study uses data from the National Educational Panel Study (NEPS), a longitudinal large-scale project on education and development from early childhood to late adulthood in Germany (Blossfeld et al., 2011). The present study focused on the starting cohort of grade 5 that consisted of regular students randomly drawn using a multi-stage stratified cluster sampling technique (starting from the year 2010/2011) to make sure that the sample is representative of students of secondary regular schools in Germany (for more details on the sampling procedure (see Steinhauer et al., 2015). The present study focused on the main sample ($n = 5,870$) and followed them at three time points: grade 5 (10 years old), grade 7 (12 years old), and grade 9 (14 years old). In the later time points, additional refreshment participants were recruited. However, these participants were not included in the present study as data were not collected from them starting from the initial wave (grade 5). Therefore, the focus of the present study is the main sample, which amounts to 52% of boys, approximately. The competence tests were administered in small groups by trained test administrators at the respective schools. The assessment was paper-based at each of the measurement points. More information on the procedures and the test and survey instruments used in starting cohort of grade 5 of the NEPS is available online (<https://www.neps-data.de/Data-Center/Data-and-Docummentation/Start-Cohort-Grade-5/Documentation>).

2.2 Measures

2.2.1 Reading Comprehension

The framework of the assessment of reading competence at the NEPS is mainly focused on the functions of text and the type of text connected to these functions and how they are related to the cognitive requirement of reading (Gehrer et al., 2013; Weinert et al., 2011). The competence was measured using five text functions namely: a) informational texts, b) commenting or argumentum texts, c) literary texts, d) instruction texts, and e) advertising reading texts (Gehrer et al. 2013). The cognitive requirements were namely: finding information text, drawing text-related conclusions, and reflecting and assessing. The approximate length of each text was from 200 to 550. Most of the tasks were designed in a multiple-choice format. The rest of the tasks were in decision making or matching format. Participants were asked whether a given statement was correct or incorrect in the decision making task while a selection of title to a corresponding text was involved in the matching task (Gehrer et al., 2013). The test duration was 28 minutes at each of the measurement points. The test consisted of a total of 32 items in grade 5, 29 items in grade 7, and 32 items in grade 9. The items were reported to support unidimensional structure. They have good item fit and high reliability score (.76 to .78), and were measurement invariant across subgroups (Pohl et al., 2012). Except for the first wave (grade 5), the tests were adapted to the ability of the participants, and respondents with low ability were given tests with relatively less difficult items. The test scores were linked across time points to make sure they are comparable over the time points for longitudinal investigation (Fischer et al., 2016). Weighted maximum likelihood estimates (WLEs) scores were used instead of sum and mean scores to take the item difficulty into account.

2.2.2 Metacognitive Monitoring

Metacognitive monitoring was assessed as a retrospective metacognitive judgement strategy (Händel et al., 2013;

Lockl, 2013), in which children reflect on the accuracy of a previous recall response (Nelson & Narens, 1994). Specifically, after completing reading tests, participants were asked to estimate their performance (“how many of the questions did you presumably answer correctly?”) (Lockl, 2013). The judgements were global as well as text specific. In the present study, we used text specific judgements. Two types of scores were calculated based on the estimation of children’s performance. The first score was the proportion of the estimated correctly solved number of items. This was calculated by dividing the number of items they judged they correctly solved by the total number of items. The second score was a deviation score, which is the difference between the proportions of correctly solved items and the proportion of the estimation of correctly solved items. For a methodological reason, instead of directly using these scores, we calculated different procedural metacognition (judgement accuracy) indexes. First, we predicted the proportion of the estimation of their performance from their actual score in reading competence. Second, we took the absolute value of the residual. We computed this score because previous studies (Dufner et al., 2015) have shown that deviation scores are problematic because they are confounded with the corresponding academic achievement to study their interrelationship. Therefore, the residual based index of procedural metacognition was used as an indicator to model procedural metacognition at a latent level. There were five indicators in grade 5 and three indicators in grades 7 and 9.

2.2.3 Socioeconomic Status

The socioeconomic status of participants was assessed by ISEI-08 (International Socio-Economic Index of Occupational Status). It is a scaling of occupations in the context of a status attainment process, i.e. how education, occupation, and earnings are obtained (Ganzeboom & Treiman, 2010). ISEI taps the features of occupation that change parents’ education into income. ISEI index is assumed to maximize the indirect effect of education on income via occupation and to minimize the direct influence of education on income (Ganzeboom et al., 1992). In other words, occupational status is assumed to convert educational credentials into earnings (Ganzeboom & Treiman, 2010). It is constructed based on the four skill level classifications of the International Standard Classification of Occupation (ISCO-08). For instance, skill level 1 involves the performance of simple and routine physical or manual tasks while skill level 4 involves tasks that require complex problem-solving, decision-making, and creativity such as professional and managerial occupations. We took the highest status in a family.

2.3 Analysis

The analysis of the study was performed using R (R Development Core Team, 2015). Structural Equation Modeling (SEM) using the package *lavaan* (Rosseel et al., 2015) in R was conducted to analyze the measurement and the main models. *comparative fit index* (CFI) and the *root mean square error of approximation* (RMSEA) values were used to evaluate the model fit: $CFI \geq .95$, $RMSEA \leq .08$ (Hu & Bentler, 1999). Specifically, confirmatory factor analysis (CFA) was employed to check the measurement models of the latent variables. In the main analysis, we used cross-lagged panel analysis (Cole & Maxwell, 2003) to test the bidirectional relationship between procedural metacognition and reading competence. We employed measurement invariance testing to make sure the latent construct, procedural metacognition, is comparable for under and over-estimating participants. A difference in the CFI of $> .01$ between two consecutive models in invariance testing (e.g., configural and weak measurement invariance models) was assumed as a serious deterioration in model fit (Cheung & Rensvold, 2002).

3. Result

3.1 Descriptive Statistics

Actual reading competence increased over time for the whole group of participants. The mean of reading comprehension (longitudinally linked WLEs scores) was $-.02$ in grade 5 and it increased to $.76$ in grade 7. In grade 9, the mean score became 1.39 . On other hand, the means of the participants’ estimation of their achievement in reading comprehension decreased from grade 5 to 7 and then it became flat from grade 7 to 9 for the participants as a whole. The mean of their estimated scores were $.73$, $.72$, and $.72$ in grades 5, 7, and 9, respectively. This implies that, on average, the participants estimated that they correctly solved 73% of the questions in grade 5 and 72% of the questions in grades 7 and 9. The absolute values of the deviation (see the formula below) between students’ estimation of their performance and their actual test score, were calculated to show the magnitude of the accuracy of judgment regardless of the direction of the self-bias for the overall group of participants. As can be seen in Table 1, the absolute deviation declined from grade 5 ($M = .22$) to 7 ($M = .21$) and 9 ($M = .18$). The decline in deviation indicated improvement in the overall monitoring accuracy over time because it showed a decline in the discrepancy between the estimation of one’s performance and the actual score.

$$Deviation = \frac{N_{estimate}}{N_{items}} - \frac{N_{correct}}{N_{items}}$$

In addition, the descriptive statistics for underestimating and overestimating groups are presented in Table 1. Participants who had a deviation score of less than zero (percentage of estimation of one's correct response in the reading minus percentage of actual score) were categorized in the underestimating group. On the other hand, participants who had a deviation score of greater than zero were categorized in the overestimating group. A deviation score of zero is assumed to be an indication of perfect judgement accuracy. The grouping was done to show the general trend as well as the difference in reading competence and monitoring accuracy scores between under- and overestimating participants. The longitudinal trend of the mean of reading competence was the same for both under and over-estimating groups and increased over time. However, the means of reading competence of underestimating group appeared to be higher than the means of the overestimating group consistently across the three-time points. The means of reading competence in the underestimating group were .83, 1.43, and 1.80 in grades 5, 7, and 9, respectively. The means of the overestimating group were $-.33$, $.28$, and 1.00 , in grades 5, 7, and 9, respectively. The mean scores of estimated scores of reading comprehension increased over time for the underestimating group while they declined for the overestimating group. As expected, the underestimating group had a lower estimation score than the overestimating group. The mean estimation scores of the underestimating group were $.64$, $.66$, and $.67$ in grades 5, 7, and 9, respectively. The means of the estimation scores of the overestimating group were $.77$, $.77$, and $.76$ at the corresponding time points. As expected, the underestimating group had a negative mean of deviation scores. Compared to the overestimating group, the underestimating group had a wider deviation in terms of magnitude. The means of the deviation scores of the underestimating group in grades 5, 7, and 9 were $-.10$, $-.12$, and $-.11$, respectively. On the other hand, the means of the deviation scores of the overestimating group were $.19$, $.19$, and $.17$ in grades 5, 7, and 9, respectively.

Table 1. Descriptive Statistics of Reading Comprehension, Estimation, and Deviation Scores

	<i>n</i>	M (SD)			Range	
		Overall	Under	Over	Min	Max
Reading grade 5	5,193	-.02 (1.27)	.82 (1.22)	-.33(1.12)	-4.71	3.96
Reading grade 7	4,044	.76 (1.36)	1.43 (1.28)	.28 (1.20)	-3.41	5.78
Reading grade 9	3,013	1.39 (1.12)	1.80 (1.06)	1.00 (1.03)	-2.12	6.24
Estimation grade 5	5,045	.73 (.16)	.64 (.18)	.77 (.14)	.00	1
Estimation grade 7	3,821	.72 (.18)	.66 (.19)	.77(.15)	.00	1
Estimation grade 9	3,106	.72 (.18)	.67(.19)	.76 (.16)	.00	1
Deviation grade 5	5,045	.22 (.12)	-.10 (.09)	.19 (.14)	.00	.85
Deviation grade 7	3,821	.21(.13)	-.12 (.12)	.19 (.15)	.00	.94
Deviation grade 9	3,106	.18 (.13)	-.11 (.11)	.17 (.15)	.00	.95

Note. Overall = the whole participants, Under = under-estimating group, Over = overestimating group; WLEs scores were used for reading comprehension; Estimation = participants' estimation of the percentage of items they correctly solved; Deviation = the actual percentage of the number of items they correctly solved minus the percentage of the estimation of the number of items they correctly solved.

Table 2. Correlation Matrix between Metacognitive Monitoring and Reading comprehension for the overall group of participants

	1	2	3	4	5
1 Monitoring grade 5					
2 Monitoring grade 7	.13				
3 Monitoring grade 9	.12	.17			
4 Reading grade 5	.31	.16	.16		
5 Reading grade 7	.24	.28	.21	.62	
6 Reading grade 9	.24	.21	.25	.61	.64

In addition, we did a correlational analysis to see the overall as well as the trend of the relationship between reading competence and procedural metacognition between under – and overestimating groups at a cross-sectional level. As the correlation matrix for the whole group of participants can be seen in Table 2, the coefficient of correlation between metacognitive monitoring and reading comprehension was moderate at the three-time points. For the overall participants, the coefficient of correlations between metacognitive monitoring and reading comprehension were .31, .28, and .25 in grades 5, 7, and 9, respectively. For the correlation as well as the cross-lagged panel analysis, we used the absolute residual scores (the procedure of the calculation is explained in the method section) to tap the monitoring accuracy of participants.

Table 3. Correlation between Metacognitive Monitoring and Reading Competence across Over- and Under-estimating Groups of Participants

	Grade 5		Grade 7		Grade 9	
	under	over	under	over	under	over
<i>r</i>	.58	.33	.46	.20	.34	.19

The relationship between metacognitive monitoring and reading comprehension seemed higher for the underestimating than the overestimating group. As can be seen in Table 3, while the correlation between procedural metacognition and reading competence at grade 5 was $r = .58$ for the underestimating group, it was $r = .33$ for the overestimating group. Similarly, in grade 7 a correlation coefficient of $r = .46$ was observed between reading comprehension and metacognitive monitoring for the underestimating group. On the other hand, the correlation coefficient was .20 for overestimating participants in grade 7. A similar trend was observed during the later grades that the correlation was higher for underestimating than the overestimating group of participants

3.2 Main Cross-Lagged Model

Before proceeding to the main analysis, we checked for the measurement model of the latent variable – metacognitive monitoring. Accordingly, we conducted a longitudinal CFA in grades 5, 7, and 9. The model fit indices of the longitudinal CFA were excellent ($n = 5,366$, $\chi^2 = 78.77$, $df = 41$, CFI = .99, RMSEA = .01, SRMR = .02). Likewise, the fit indices of the cross-lagged model for the whole group of participants were good ($n = 5,754$, $\chi^2 = 503.30$, $df = 105$, CFI = .95, RMSEA = .03, SRMR = .03). In the main cross-lagged model we controlled for the SES of the participants because previous research indicated the positive role of SES on both metacognition and achievement (Thompson and Foster 2014; Yerdelen-Damar and Peşman 2013). The cross-lagged model was mainly analyzed to see whether metacognitive monitoring and reading competence had a bidirectional relationship between grades 5 to 7 and grades 7 to 9. To control for the previous metacognitive monitoring and reading comprehension, the autoregressive effects were taken into account. While metacognitive monitoring was computed at a latent level, longitudinally linked WLEs scores of reading comprehension were used in the cross-lagged panel analysis. The missing values were handled using a full maximum likelihood information (FIML) function.

For the overall participants, we found low developmental stability in metacognitive monitoring from grade 5 to 7 ($\beta = .20$, $p < .01$) but a slight increment was observed in the later grades from 7 to 9 ($\beta = .30$, $p < .01$). On the other hand, a relatively moderate developmental stability in reading comprehension was observed from grade 5 to 7 ($\beta = .54$, $p < .01$) and 7 to 9 ($\beta = .56$, $p < .01$) as depicted in Figure 1. The cross-lagged panel analysis result revealed that the development of reading comprehension at grade 7 was positively predicted ($\beta = .09$, $p < .01$) by procedural metacognition at grade 5 after controlling for reading comprehension at grade 5 (see Figure 1). Similarly, a relatively smaller but significant cross-lagged effect from metacognitive monitoring to reading competence was consistently shown in the later grades. Metacognitive monitoring grade 7 positively and significantly predicted the development of reading comprehension at grade 9 ($\beta = .07$, $p < .01$) after controlling for reading comprehension at grade 7.

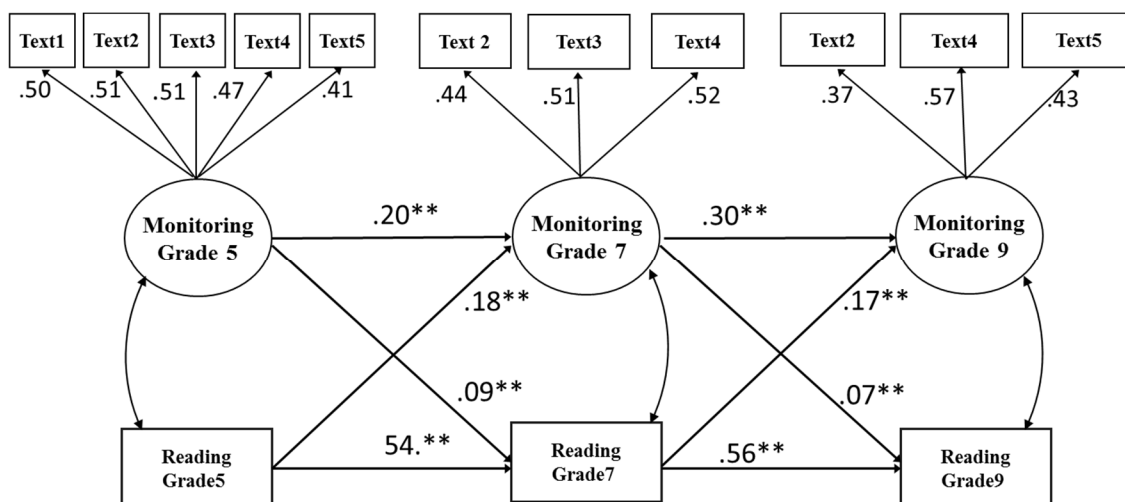


Figure 1. Main Cross-lagged Model (Overall group of Participants)

Note. $n = 5,754$, $\chi^2 = 503.30$, $df = 105$, $CFI = .95$, $RMSEA = .03$, $SRMR = .03$. The socioeconomic status of participants was controlled. All of the coefficients are standardized. $P^* < .05$, $P^{**} < .01$

In the opposite direction, stronger effects from reading comprehension to metacognitive monitoring were observed. Reading comprehension in grade 5 had a positive and significant developmental effect on metacognitive monitoring at grade 7 ($\beta = .18$, $p < .01$) after controlling for metacognitive monitoring in grade 5 (see Figure 1). In the later grades, the development of metacognitive monitoring at grade 9 was significantly and positively predicted by reading comprehension at grade 7 ($\beta = .17$, $p < .01$) after controlling for metacognitive monitoring at grade 7 as depicted in Figure 1. Overall, the cross-lagged panel analysis showed the developmental relationship between reading competence and metacognitive monitoring to be reciprocal. However, there was a consistent difference in the strength of the effects that reading comprehension had more effect than metacognitive monitoring. This difference was statistically confirmed using a χ^2 -difference test that evaluated the model fit difference between a restricted (i.e., where the reciprocal effects were equal) and an unrestricted model. The χ^2 -difference test showed the effect of reading comprehension to be significantly greater than the effect of metacognitive monitoring from grade 5 to 7 ($\Delta\chi^2(1, N = 5,754) = 13.75$, $p < .001$) and from grade 7 to 9 ($\Delta\chi^2(1, N = 5,754) = 5.96$, $p < .001$).

3.3 Multi-Group Cross-Lagged Models (Under - vs Overestimation)

To further understand the trend of the longitudinal relationship between reading comprehension and metacognitive monitoring between under and overestimating participants, a multi-group cross-lagged panel analysis was conducted. We could not compute a multi-group model including the three time points since the group members were dynamic over the period. If the group variable was constant (such as sex), we could have computed the multi-group cross-lagged models for the three-time points in tandem. Therefore, we divided the time points into two. First, we conducted the multi-group cross-lagged model for the time points grade 5 and 7 between those participants who under – and overestimated their performance in grade 5. Consequently, we did the same analysis for the time points in grades 7 and 9, between those participants who under – and overestimated their performance in reading in grade 7.

Before computing the multi-group cross-lagged models, we checked for the measurement invariance of the latent variable – metacognitive monitoring. For the first grouping (between under – and overestimating groups at grade 5), the model fit indices of the configural measurement invariance of the procedural metacognition were acceptable ($\chi^2 = 53.16$, $df = 38$, $p < .001$; $CFI = .99$; $RMSEA = .01$), which suggested similar factor structure between the two groups (see Table 4.). In the next step, weak measurement invariance testing, the factor loadings were constrained to be equal between under and overestimating groups and it did not show meaningful model fit deterioration ($\chi^2 = 70.45$, $df = 44$, $p < .001$; $\Delta CFI = .00$) as ΔCFI was not above the cutoff point (.01) suggested by Cheung and Rensvold (2002). In the third step, testing for strong measurement invariance, the factor loadings, and intercepts were constrained to be equal across the two groups. The fit indices did not exhibit meaning model fit deterioration ($\chi^2 = 96.54$, $df = 50$, $p < .001$; $\Delta CFI = .01$) because ΔCFI did not exceed the threshold. However, in the strict invariance (factor loadings, intercepts, and residuals constrained to be equal) the model fit showed significant deterioration ($\chi^2 = 399.99$, $df = 58$, $p < .001$; $\Delta CFI = .15$) because ΔCFI was above the cutoff point, .01.

Table 4. Metacognitive Monitoring Measurement Invariance Testing

	Group grade 5				Group grade 7			
	χ^2/df	CFI	RMSEA	$\Delta CFI/df$	χ^2/df	CFI	RMSEA	$\Delta CFI/df$
Configural invariance	53.16/38	.99	.01		19.62/16	.99	.01	
Weak invariance	70.45/44	.99	.02	.00/6	34.63/20	.98	.02	.01/4
Strong invariance	96.54/50	.98	.02	.01/6	35.37/24	.99	.01	-.01/4
Strict invariance	399.99/58	.83	.05	.15/8	144.53/30	.89	.04	.10/6

Note. Configural = similar factor structure; Weak = factor loadings are constrained to equality; Strong = intercepts are constrained to equality in addition to factor loadings; Strict = residuals are constrained to equality in addition to factor loadings and intercepts.

We computed similar measurement invariance testing for metacognitive monitoring under – and overestimating groups at grade 7 and the change in the model fit indices confirmed strong measurement invariance as presented in Table 4. Therefore, strong measurement invariance was imposed in the subsequent multi-group cross-lagged panel analysis models. This would exclude the assumption that the difference observed between the two groups was merely because of a difference in the meaning of the measure between under – and overestimating groups of participants.

The fit indices of the multi-group cross-lagged models (Figures 2 and 3) based on the grouping at grade 5 ($\chi^2 = 293.45$, $df = 88$, CFI = .96, RMSEA = .03, SRMR = .04) and grade 7 ($\chi^2 = 155.40$, $df = 50$, CFI = .97, RMSEA = .03, SRMR = .04) were acceptable. For the underestimating group, there was a significant and positive reciprocal effect between reading comprehension and metacognitive monitoring from grade 5 to 7 taking the prior relationship into account (see Figure 2). Metacognitive monitoring at grade 5 had a positive and significant effect on the development of reading comprehension at grade 7 ($\beta = .21$, $p < .01$) taking reading competence at grade 5 into account for the underestimating group of participants. Similarly, reading comprehension in grade 5 had a positive and significant effect ($\beta = .15$, $p < .05$) on the development of metacognitive monitoring in grade 7 controlling for metacognitive monitoring in grade 5. Similarly, there was a reciprocal developmental relationship between metacognitive monitoring and reading comprehension for overestimating group from grade 5 to 7 although the magnitudes of the effects were lower compared to the underestimating group. Metacognitive monitoring at grade 5 had a positive and significant developmental effect ($\beta = .07$, $p < .05$) on reading comprehension in grade 7 after controlling for reading comprehension at grade 5. On the other direction, there was a positive and significant effect ($\beta = .11$, $p < .05$) from reading comprehension at grade 5 to metacognitive monitoring at grade 7 after controlling for metacognitive monitoring at grade 5.

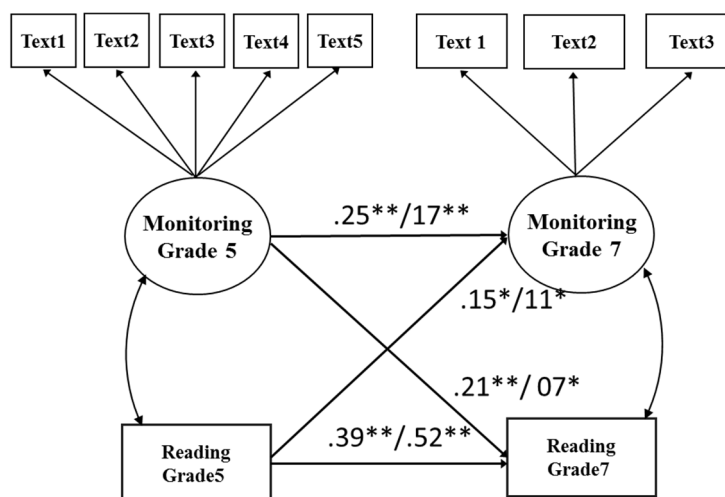


Figure 2. Multi-Group Cross-Lagged Model (Grade 5)

Note. The groups were those who underestimated and overestimated their performance at grade 5; n (under) = 1,443 n (over) = 3,602, $\chi^2 = 293.45$, $df = 88$, CFI = .96, RMSEA = .03, SRMR = .04; The parameters on the left side

are for underestimating the group of participants; Strong measurement invariance was imposed between the groups; SES of the participants was controlled; $P^* < .05$, $P^{**} < .01$.

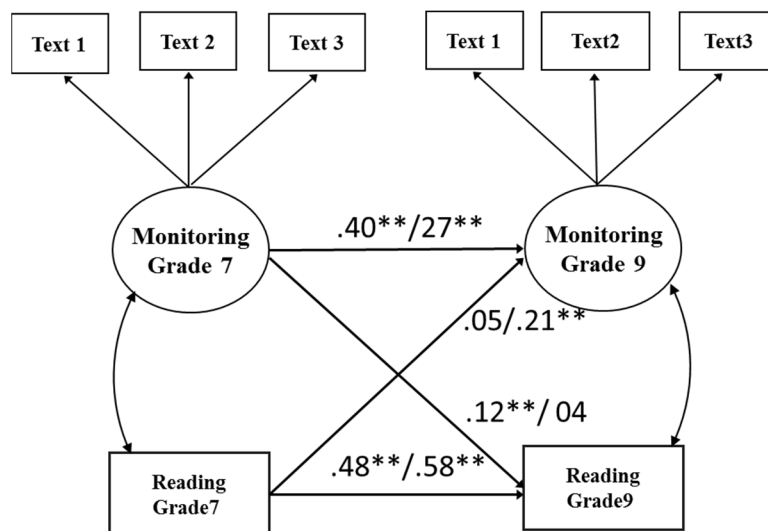


Figure 3. Multi-Group Cross-Lagged Model (Grade 7)

Note. The groups were those who underestimated and overestimated their performance at grade 7; n (under) = 1,680 n (over) = 2, 441 $\chi^2 = 155.40$, $df = 40$, CFI = .97, RMSEA = .03, SRMR = .04; The parameters on the left side are for underestimating the group of participants; Strong measurement invariance was imposed between the groups; The socioeconomic status of participants was controlled; $P^* < .05$, $P^{**} < .01$.

However, a difference in the developmental relationship between metacognitive monitoring and reading comprehension was observed between under – and overestimating groups at the later time points, from grade 7 to 9. The developmental relationship became uni-directional for both groups but in a different direction. While a positive and significant effect ($\beta = .12$, $p < .01$) from metacognitive monitoring in grade 7 to reading comprehension in grade 9 was evident for underestimating group, in the opposite direction, there was a significant effect ($\beta = .21$, $p < .01$) from reading comprehension at grade 7 to metacognitive monitoring at grade 9 for overestimating group. The developmental effects from reading comprehension at grade 7 to metacognitive monitoring at grade 9, and metacognitive monitoring at grade 7 to reading comprehension at grade 9 became insignificant for under and over-estimating groups, respectively, as depicted in Figure 3.

4. Discussion

The study intended to examine whether the relationship between metacognitive monitoring and reading comprehension is reciprocal. The monitoring aspect of procedural metacognition, which was measured by the accuracy of self-judgment of performance in reading comprehension, was the particular focus of the study. The finding of the study has confirmed the hypothesis that metacognitive monitoring and reading comprehension has a reciprocal developmental relationship from grade 5 to 7 and grade 7 to 9. The observed positive effects of metacognitive monitoring on reading comprehension support the established literature on the importance of procedural metacognition on academic achievement in general and reading competence in particular (Pintrich, 2002; Schneider, 2015). From a metacognition research perspective, it can be assumed that monitoring is useful. It provides accurate information about the current epistemic state to regulate learning processes (Nelson & Narens, 1994). And this information, obtained through monitoring, helps students to properly allocate their time during their studies and employ appropriate learning strategies (Thiede & Dunlosky, 1999). This illustrates the importance of monitoring accuracy in providing ongoing feedback on one's learning process to initiate self-regulation such as whether to terminate or prolong study time (Hacker et al., 2000; Thiede & Dunlosky, 1999). Previous empirical investigations also demonstrated consistent results with the present study that metacognitive monitoring accuracy plays a positive role in reading comprehension (Bouffard et al., 1998; Roebers et al., 2014; Stankov et al., 2012).

Overall, students' metacognitive monitoring accuracy had positive and significant developmental effects on reading comprehension from grade 5 to 7 and grade 7 to 9. However, reading comprehension had stronger

reciprocal effects on later metacognitive monitoring accuracy from grade 5 to 7 and grade 7 to 9. The reciprocal effects were small but it should be noted that autoregressive effects and SES were controlled. This longitudinal finding is consistent with the general theoretical assumption that metacognition and academic achievement might be mutually interdependent throughout development (Flavell & Wellman, 1977; Schneider, 1985; Schneider & Pressley, 1997). In a specific focus on metacognitive monitoring, the skill required to accurately judge one's performance in reading is similar to the skill required to effectively comprehend a given text (Dunning et al., 2003). Therefore, low achievers might be in a double curse as they fail to judge their performance the way they fail to comprehend a given text. (Dunning et al., 2003; Miller & Geraci, 2011). This indicates the role of knowledge in a subject matter on the later development of monitoring accuracy. Among a few previous studies that focus on the reciprocal relationship between procedural metacognition and reading comprehension, the findings of the present study appear to be partially inconsistent with a recent study conducted on secondary school students by King and McInerney (2016). Instead of a reciprocal relationship, they found a uni-directional effect from academic achievement (mathematics and reading comprehension) to procedural metacognition (monitoring and planning). However, it should be noted that there was a difference in the measure between theirs and the present study that they used a self-report measure to assess metacognitive ability.

In addition to the overall investigation of the developmental relationship between metacognitive monitoring and reading comprehension, the study aimed to examine whether the result observed in the overall group of participants holds to be true for both under – and overestimating groups. The multi-group cross-lagged panel analysis indicated that metacognitive monitoring and reading comprehension had a reciprocal developmental relationship from grades 5 to 7. That means monitoring accuracy and reading comprehension were mutually interdependent for both groups despite the strength of the association being more pronounced for the underestimating group. However, the relationship became uni-directional from grade 7 to 9 and the effects were from metacognitive monitoring to reading achievement for the underestimating group, and from reading to metacognitive monitoring for the overestimating group. This could be because the overestimating group had lower reading achievement (as was observed in the descriptive statistics in Table 1) at the initial time point and its development could help the students to improve their monitoring accuracy at grade 9. On the other hand, underestimating students were already competent in reading comprehension during the initial period thus effectively monitoring their learning process could further contribute to the development of reading comprehension.

In addition, the association between metacognitive monitoring and reading comprehension was more pronounced for the underestimating group in the cross-sectional correlation analysis at the three time points. When the multi-group cross-lagged panel analysis is combined with the cross-sectional correlation analysis, monitoring accuracy appeared to be useful for the later development of reading comprehension. However, judgement accuracy seemed more important for those students who underestimated their competence than overestimating students. This was evident even after imposing strong measurement invariance between the two groups in the longitudinal analysis. On the other hand, the means of the reading comprehension of the underestimating group were higher than the means of the overestimating group. On contrary, the means of the deviation of the underestimating group were less than the overestimating group across the three-time points. Therefore, the observed higher means of the underestimating group in reading comprehension are not surprising given the means of the estimation scores of the underestimating group were closer to their actual scores than the means of the estimation scores of the overestimating group – as accuracy leads to high achievement. This could also mean excessive overestimation is more hindering than moderate underestimation. Self-handicapping tendency might also be another explanation for why the underestimating group had higher achievement scores in reading than the overestimating group. Some students might underestimate their achievements to protect their self-esteem (Elliot & Church, 2003). On the other hand, the fact that accuracy and achievement were more associated with underestimation than overestimation – in the longitudinal analysis – might imply that excessive underestimation may impair the development of reading comprehension to a greater extent than excessive overestimation. This could be because excessive underestimation might create negative emotions that limit motivation (Bouffard & Narciss, 2011).

To summarize, the study illustrates the importance of metacognitive monitoring to provide the right ongoing feedback to regulate learning processes as required (Hacker et al., 2000; Thiede & Dunlosky, 1999). Low metacognitive accuracy could be because of either under – or overestimation of one's performance. Consistent underestimation of one's performance might have a negative impact on reading comprehension because it affects motivation (Bouffard et al., 2011; Efklides, 2006). From the perspective of metacognition research, overestimation is also assumed to have a negative impact on academic achievement as it might be an obstacle to identifying one's own learning needs and this might make students less prepared and less interested in seeking help from others

(Butler & Winne, 1995; Dunlosky et al., 2005; Stone & May, 2002). While the result of the study suggested the benefit of the accuracy of one's judgement for reading comprehension, comparatively, excessive underestimation tends to be more hindering than excessive overestimation for later achievement in reading comprehension. The result of the present study has shown this complex longitudinal relationship using large-scale data and a measure that did not depend on social comparison but was based on the residual of a criterion (actual test score), which is less confounded to achievement. Often deviation scores between self-evaluation and actual score are used to assess monitoring accuracy and this was reported to be problematic because it is highly confounded with achievement (Dufner et al., 2015). To overcome this problem, using the residual after predicting the criterion on self-evaluation is recommended (Dufner et al., 2015; Gonida & Leondari, 2011). The residual based score strengthens the assumption the relationship observed between metacognitive monitoring and reading comprehension was not because the measures were simply related technically. And this is believed to be an important contribution to shedding a light on the debate on the developmental relationship between metacognitive monitoring and reading comprehension.

5. Conclusion

The findings of the study are valuable to contribute to the scientific debate on the longitudinal relationship between metacognitive monitoring and reading comprehension. It demonstrated the developmental interplay between the monitoring aspect of metacognition and reading comprehension in large-scale data. It further provided longitudinal insights into the implication of under – and overestimation of one's performance on later reading comprehension. These have practical implications such that promoting students to accurately monitor their own learning process could have a positive effect on academic achievement in general and reading comprehension in particular. Discouraging under – and overestimation of one's own performance might pave the way to effectively regulate the learning process. Therefore, parents and teachers are recommended to help children develop a realistic view of their performance. One of the mechanisms of fostering students' monitoring ability can be encouraging the competence of children in the subject matter as metacognitive monitoring and achievement are mutually interdependent in the course of development. Despite these theoretical and practical implications, it is worth mentioning one of the limitations of this study i.e. only one aspect of procedural metacognition was investigated. Therefore, it may be fruitful to investigate the developmental interplay by including other aspects of procedural metacognition such as the allocation of study time. In addition, it may also be worthwhile to examine the relationships with shorter time intervals between the measurement points (such as micro genetic study) to analyze how monitoring accuracy affects metacognitive control and learning behavior immediately. Moreover, examining whether the observed reciprocal relationship holds to be true with other domains of academic achievement might be the focus of future research.

Acknowledgments

This work was supported by the Bamberg Graduate School of Social Sciences which is funded by the German Research Foundation (DFG) under the German Excellence Initiative (GSC 1024).

References

- Artelt, C., & Schneider, W. (2015). Cross-Country Generalizability of the Role of Metacognitive Knowledge in Students' Strategy Use and Reading Competence. *Teachers College Record, 117*(1), n1. <https://doi.org/10.1177/016146811511700104>
- Bjorklund, D. F., & Bering, J. M. (2002). The evolved child. *Learning and Individual Differences, 12*(4), 347-373. [https://doi.org/10.1016/S1041-6080\(02\)00047-X](https://doi.org/10.1016/S1041-6080(02)00047-X)
- Blossfeld, H.-P., Roßbach, H.-G., von Maurice, J., Schneider, T., Kiesl, S. K., Schönberger, B., ... Prenzel, M. S. (2011). Education as a Lifelong Process-The german national educational Panel study (nePs). *Age, 74*(73), 72.
- Bouffard, T., & Narciss, S. (2011). Benefits and risks of positive biases in self-evaluation of academic competence: Introduction. *International Journal of Educational Research, 50*(4), 205-208. <https://doi.org/10.1016/j.ijer.2011.08.001>
- Bouffard, T., Cote, S., Larouche, M., Vaillancourt, M., & Fleury-Roy, M. (2006). Effects of positive illusory biases among elementary school children. *Presented at the 10th international conference on motivation.*
- Bouffard, T., Markovits, H., Vezeau, C., Boisvert, M., & Dumas, C. (1998). The relation between accuracy of self-perception and cognitive development. *British Journal of Educational Psychology, 68*(3), 321-330. <https://doi.org/10.1111/j.2044-8279.1998.tb01294.x>

- Bouffard, T., Vezeau, C., Roy, M., & Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research*, 50(4), 221-229. <https://doi.org/10.1016/j.ijer.2011.08.003>
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281. <https://doi.org/10.3102/00346543065003245>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5
- Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, 20(1), 2-17. <https://doi.org/10.1016/j.learninstruc.2008.11.002>
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, 112(4), 558-577. <https://doi.org/10.1037/0021-843X.112.4.558>
- Dufner, M., Reitz, A. K., & Zander, L. (2015). Antecedents, Consequences, and Mechanisms: On the Longitudinal Interplay Between Academic Self-Enhancement and Psychological Adjustment: Academic Self-Enhancement and Psychological Adjustment. *Journal of Personality*, 83(5), 511-522. <https://doi.org/10.1111/jopy.12128>
- Dunlosky, J., Hertzog, C., Kennedy, M. R., & Thiede, K. W. (2005). *The Self-Monitoring Approach For Effective Learning*. Cognitive Technology.
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12(3), 83-87. <https://doi.org/10.1111/1467-8721.01235>
- Edossa, A. K., Neuenhaus, N., Artelt, C., Lingel, K., & Schneider, W. (2019). Developmental relationship between declarative metacognitive knowledge and reading comprehension during secondary school. *European Journal of Psychology of Education*, 34(2), 397-416. <https://doi.org/10.1007/s10212-018-0393-x>
- Edossa, A. K., Schroeders, U., Weinert, S., & Artelt, C. (2017). The development of emotional and behavioral self-regulation and their effects on academic achievement in childhood. *International Journal of Behavioral Development*. <https://doi.org/10.1177/0165025416687412>
- Efkliades, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1(1), 3-14. <https://doi.org/10.1016/j.edurev.2005.11.001>
- Elliot, A. J., & Church, M. A. (2003). A motivational analysis of defensive pessimism and self-handicapping. *Journal of Personality*, 71(3), 369-396. <https://doi.org/10.1111/1467-6494.7103005>
- Fischer, L., Rohm, T., Gnamb, T., & Carstensen, C. (2016). *Linking the data of the competence tests (NEPS Survey Paper No. 1)*. Bamberg: Leibniz Institute for Educational Trajectories, National Education Panel Study.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906. <https://doi.org/10.1037/0003-066X.34.10.906>
- Flavell, J. H., & Miller, P. H. (2002). *Cognitive development* (4th ed). Upper Saddle River, N.J: Prentice Hall.
- Fritz, K., Howie, P., & Kleitman, S. (2010). "How do I remember when I got my dog?" The structure and development of children's metamemory. *Metacognition and Learning*, 5(2), 207-228. <https://doi.org/10.1007/s11409-010-9058-0>
- Ganzeboom, H. B., & Treiman, D. J. (2010). Occupational status measures for the new International Standard Classification of Occupations ISCO-08; with a discussion of the new classification. Presented at the *Annual Conference of International Social Survey Programme*, Lisbon.
- Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1-56. [https://doi.org/10.1016/0049-089X\(92\)90017-B](https://doi.org/10.1016/0049-089X(92)90017-B)
- Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study/NEPS-Rahmenkonzeption zur Messung von Lesekompetenz und Resultate einer Pilotstudie mit Erwachsenen. *Journal for Educational Research Online*, 5, 50.
- Gonida, E. N., & Leondari, A. (2011). Patterns of motivation among adolescents with biased and accurate

- self-efficacy beliefs. *International Journal of Educational Research*, 50(4), 209-220. <https://doi.org/10.1016/j.ijer.2011.08.002>
- 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. <https://doi.org/10.1037/0022-0663.92.1.160>
- Händel, M., Artelt, C., & Weinert, S. (2013). Assessing metacognitive knowledge: Development and evaluation of a test instrument. *Journal for Educational Research Online*, 5, 162-180. <https://doi.org/10.1037/t67625-000>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Huff, J. D., & Nietfeld, J. L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning*, 4, 161-176. <https://doi.org/10.1007/s11409-009-9042-8>
- King, R. B., & McInerney, D. M. (2016). Do goals lead to outcomes or can it be the other way around?: Causal ordering of mastery goals, metacognitive strategies, and achievement. *British Journal of Educational Psychology*, 86(2), 296-312. <https://doi.org/10.1111/bjep.12107>
- Koyuncu, I., Bulus, M., & Firat, T. (2021). The Moderator Role of Gender and Socioeconomic Status in the Relationship Between Metacognitive Skills and Reading Scores. *Participatory Educational Research*, 9(3), 82-97. <https://doi.org/10.17275/per.22.55.9.3>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Kwan, V. S., John, O. P., Robins, R. W., & Kuang, L. L. (2008). Conceptualizing and assessing self-enhancement bias: a componential approach. *Journal of Personality and Social Psychology*, 94(6), 1062. <https://doi.org/10.1037/0022-3514.94.6.1062>
- Liu, J., Peng, P., & Luo, L. (2020). The Relation Between Family Socioeconomic Status and Academic Achievement in China: A Meta-analysis. *Educ Psychol Rev*, 32, 49-76. <https://doi.org/10.1007/s10648-019-09494-0>
- Lockl, K. (2013). *Assessment of procedural metacognition: Scientific Use File 2013*. Bamberg: University of Bamberg, National Educational Panel Study.
- Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), 502-506. <https://doi.org/10.1037/a0021802>
- Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. In *Psychology of Learning and Motivation* (Vol. 26, pp. 125-173). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)60053-5](https://doi.org/10.1016/S0079-7421(08)60053-5)
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition?
- Paris, S. G. (2002). When is metacognition helpful, debilitating, or benign? In *Metacognition* (pp. 105-120). Springer. https://doi.org/10.1007/978-1-4615-1099-4_8
- Paris, S. G., & Byrnes, J. P. (1989). The constructivist approach to self-regulation and learning in the classroom. In *Self-regulated learning and academic achievement* (pp. 169-200). Springer. https://doi.org/10.1007/978-1-4612-3618-4_7
- Paris, S. G., Wasik, B., & Turner, J. C. (1991). *The development of strategic readers*.
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, 41, 219-225. https://doi.org/10.1207/s15430421tip4104_3
- Pohl, S., Haberkorn, K., Hardt, K., & Wiegand, E. (2012). *NEPS Technical Report for reading-Scaling results of starting cohort 3 in fifth grade*. NEPS Working Paper.
- R Development Core Team. (2016). *R: A language and environment for statistical computing* (Version 2.13.2). Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Roderer, T., & Roebbers, C. M. (2010). Explicit and implicit confidence judgments and developmental differences in metamemory: an eye-tracking approach. *Metacognition and Learning*, 5(3), 229-250.

- <https://doi.org/10.1007/s11409-010-9059-z>
- Roebbers, C. M., Cimeli, P., Röthlisberger, M., & Neuenschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: an explorative study on their interrelations and their role for school achievement. *Metacognition and Learning*, 7(3), 151-173. <https://doi.org/10.1007/s11409-012-9089-9>
- Roebbers, C. M., Krebs, S. S., & Roderer, T. (2014). Metacognitive monitoring and control in elementary school children: Their interrelations and their role for test performance. *Learning and Individual Differences*, 29, 141-149. <https://doi.org/10.1016/j.lindif.2012.12.003>
- Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., ... Barendse, M. (2015). Package 'lavaan.'
- Schneider, W. (1985). Developmental trends in the metamemory-memory behavior relationship: an integrative review. In D.-L. Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (pp. 57-109). Orlando: Academic Press.
- Schneider, W. (2015). The development of metamemory. In *Memory development from early childhood through emerging adulthood* (pp. 255-308). New York: Springer. https://doi.org/10.1007/978-3-319-09611-7_9
- Schneider, W., & Artelt, C. (2010). Metacognition and mathematics education. *ZDM*, 42(2), 149-161. <https://doi.org/10.1007/s11858-010-0240-2>
- Schneider, W., & Lockl, K. (n.d.). Procedural Metacognition in Children: Evidence for Developmental Trends. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 391-410). New York: Taylor & Francis Group, LLC.
- Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty* (2nd ed.). Mahwah: Erlbaum.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7, 351-371. <https://doi.org/10.1007/BF02212307>
- Shin, H., Bjorklund, D. F., & Beck, E. F. (2007). The adaptive nature of children's overestimation in a strategic memory task. *Cognitive Development*, 22(2), 197-212. <https://doi.org/10.1016/j.cogdev.2006.10.001>
- Soto, C., Gutierrez de Blume, A. P., Carrasco Bernal, M. A., & Contreras Castro, M. A. (2020). The role of metacognitive cues on the comprehension of proficient and poor readers. *Journal of Research in Reading*, 43(3), 272-289. <https://doi.org/10.1111/1467-9817.12303>
- 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(6), 747-758. <https://doi.org/10.1016/j.lindif.2012.05.013>
- Steinhauer, H. W., Aßmann, C., Zinn, S., Goßmann, S., & Rässler, S. (2015). Sampling and weighting cohort samples in institutional contexts. *AStA Wirtschafts - und Sozialstatistisches Archiv*, 9, 131-157. <https://doi.org/10.1007/s11943-015-0162-0>
- Stone, C. A., & May, A. L. (2002). The accuracy of academic self-evaluations in adolescents with learning disabilities. *Journal of Learning Disabilities*, 35(4), 370-383. <https://doi.org/10.1177/00222194020350040801>
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 1024. <https://doi.org/10.1037/0278-7393.25.4.1024>
- Thiede, K. W., Anderson, M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95(1), 66. <https://doi.org/10.1037/0022-0663.95.1.66>
- Thompson, R. B., & Foster, B. J. (2014). Socioeconomic Status and Parent-Child Relationships Predict Metacognitive Questions to Preschoolers. *Journal of Psycholinguistic Research*, 43(4), 315-333. <https://doi.org/10.1007/s10936-013-9256-4>
- Yerdelen-Damar, S., & Peşman, H. (2013). Relations of gender and socioeconomic status to physics through metacognition and self-efficacy. *The Journal of Educational Research*, 106(4), 280-289. <https://doi.org/10.1080/00220671.2012.692729>
- Zargar, E., Adams, A. M., & Connor, C. M. (2020). The relations between children's comprehension monitoring and their reading comprehension and vocabulary knowledge: an eye-movement study. *Reading and writing*,

33(3), 511-545. <https://doi.org/10.1007/s11145-019-09966-3>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).