The Double Curse of Misconceptions: Misconceptions Impair not Only Text Comprehension
but Also Metacomprehension in the Domain of Statistics

Anja Prinz, Stefanie Golke, and Jörg Wittwer
University of Freiburg

Author Note

Anja Prinz, Department of Educational Science, University of Freiburg, ORCID:
0000-0002-1097-3442; Stefanie Golke, Department of Educational Science, University of
Freiburg, ORCID: 0000-0002-0227-9303; Jörg Wittwer, Department of Educational Science,
University of Freiburg.

This study was funded by the German Federal Ministry of Education and Research
(BMBF; 01JA1518A). The authors declare that they have no conflict of interest.

Correspondence concerning this article should be addressed to Anja Prinz,
Department of Educational Science, University of Freiburg, Rempartstr. 11, 79098 Freiburg,
Germany, E-mail: anja.prinz@ezw.uni-freiburg.de, Tel. +49-761-2039130

Manuscript version of:
Prinz, A., Golke, S., & Wittwer, J. (2018). The double curse of misconceptions: Misconceptions impair not only text comprehension but also metacomprehension in the domain of statistics. *Instructional Science*. Advance online publication. doi:10.1007/s11251-018-9452-6
The final publication is available at http://dx.doi.org/10.1007/s11251-018-9452-6
Abstract

Research shows that misconceptions are usually detrimental to text comprehension. However, whether misconceptions also impair metacomprehension accuracy, that is, the accuracy with which one self-assesses one’s text comprehension, has received far less attention. We conducted a study in which we examined students’ \((N = 47)\) comprehension and metacomprehension accuracy (prediction accuracy and postdiction accuracy) of a statistics text as a function of their statistical misconceptions. Text comprehension and metacomprehension accuracy referred to both conceptual and procedural aspects of statistics. The results showed that students who had more misconceptions achieved poorer conceptual text comprehension and, at the same time, provided more overconfident predictions of their conceptual and procedural text comprehension than students who had fewer misconceptions. In contrast, postdiction accuracy of conceptual and procedural text comprehension was not affected by misconceptions.

Keywords: conceptual and procedural knowledge, metacomprehension accuracy, misconceptions, statistics, text comprehension
The Double Curse of Misconceptions: Misconceptions Impair not Only Text Comprehension but Also Metacomprehension in the Domain of Statistics

Nothing is more dangerous for a new truth than an old misconception.

—Johann Wolfgang von Goethe, *Wilhelm Meister's Journeyman Years, 1829*

Research in many domains has revealed that unskilled people often not only show poor performance but also lack awareness of their poor performance (e.g., Dunning et al. 2003; Kruger and Dunning 1999). This phenomenon, known as the double curse of incompetence (Dunning et al. 2003), has also been found with regard to text comprehension. More specifically, learners who have little prior knowledge about the content of a text normally have difficulty to comprehend the text and, at the same time, to accurately self-assess their comprehension (known as metacomprehension accuracy; Griffin et al. 2009; Schneider et al. 1989). So far research, however, has given relatively little attention to the question whether text comprehension and metacomprehension accuracy are impaired not only when learners have little prior knowledge but also when learners have flawed prior knowledge in form of misconceptions (for exceptions concerning text comprehension, see, e.g., Kendeou and van den Broek 2005). We conducted an empirical study to examine the effects of learners’ misconceptions on their text comprehension and metacomprehension accuracy. In doing so, we focused on misconceptions in the domain of statistics.

**Misconceptions in the Domain of Statistics**

The acquisition of statistical skills is highly relevant in today’s information societies (Gal 2002). Therefore, statistics has become a compulsory part of many university programs, for example in the social sciences, with the aim to enable students to comprehend, handle, and produce scientific information (Ben-Zvi and Garfield 2004). Learning statistics not only
means to know how to apply statistical procedures but also to understand the concepts and principles underlying the statistical procedures (Ben-Zvi and Garfield 2004; Garfield and Ben-Zvi 2007). For example, when analyzing the covariance between two variables, it is important for learners to be able to calculate the covariance based on the covariance formula as well as to understand what covariance actually means. Thus, procedures and concepts are of equal importance in learning statistics, which is in accordance with the widely accepted view that procedural knowledge and conceptual knowledge of mathematics represent two different types of knowledge that are, however, highly associated with each other (e.g., Rittle-Johnson and Schneider 2015). For example, Schneider et al. (2011) not only revealed a substantial correlation between procedural and conceptual knowledge but also showed that procedural knowledge supported the acquisition of conceptual knowledge and vice versa.

However, many learners have difficulties in learning statistical concepts and procedures (Garfield and Ahlgren 1988). One major source of these difficulties lies in the fact that learners often have misconceptions about statistical concepts (see, e.g., Huck 2016). Such misconceptions usually differ in fundamental ways from the normatively correct concepts (Liu et al. 2010). Well-known examples are statistical misconceptions about measures of association such as covariance (e.g., Liu et al. 2009; Moritz 2004). For instance, learners often erroneously assume that covariance between two variables proves a causal relationship, which can impair the correct interpretation of the association between two variables (e.g., Morris 1999). In addition, the misconception that a straight fit line invariably indicates perfect covariance might lead to the incorrect conclusion that even a null variance with a straight fit line reflects perfect covariance (e.g., Liu 2010). These examples show that misconceptions can seriously hamper the acquisition of concepts and procedures in learning statistics.
Misconceptions and Text Comprehension

Successful comprehension of a text requires that learners actively process the information provided in the text and construct a mental representation of this information. To do so, they need to integrate the textual information with each other and with their prior knowledge (Kintsch 1998). However, when learners have flawed prior knowledge in form of misconceptions, comprehension of a text can be seriously hampered (e.g., Alvermann et al. 1985; Diakidoy and Kendeou 2001; see also Guzzetti et al. 1993; Tippett 2010). Research suggests that learners who have misconceptions remain fairly unaware of conflicts between their flawed knowledge and the information provided in a text (see, e.g., Guzzetti et al. 1993; Tippett 2010). For example, Kendeou and van den Broek (2005, 2007) examined the online processes taking place when learners with misconceptions read a text. Their results showed that learners with misconceptions engaged in the same types of reading processes, for example, inferring, elaborating, and explaining, as learners without misconceptions. However, the content of their reading processes was contaminated by their misconceptions. Therefore, their mental representation of the text was flawed and comprehension was impaired.

Misconceptions and Metacomprehension Accuracy

To effectively learn from reading a text, it is important that learners not only construct an appropriate mental representation of the text but also accurately monitor and judge their comprehension of the text, which is known as metacomprehension accuracy (e.g., Thiede et al. 2009). For example, if learners accurately judge which text material they have not comprehended well yet, they can devote their attention to these unlearned parts. In contrast, if metacomprehension accuracy is poor, learners will not be able to appropriately self-regulate their learning (e.g., Thiede et al. 2003). Previous studies show that the amount of a learner’s prior knowledge affects metacomprehension accuracy. For example, Griffin et al. (2009)
found that learners with less prior knowledge about baseball not only achieved poorer comprehension on a set of texts on baseball topics but were also less accurate in predicting their comprehension than learners with more prior knowledge about baseball (for similar results, see Schneider et al. 1989).

However, the role of flawed prior knowledge in form of misconceptions for metacomprehension accuracy has received little attention so far. To understand how misconceptions might influence metacomprehension accuracy, the cue-utilization framework provides a useful theoretical basis (Griffin et al. 2009; cf. Koriat 1997). An important distinction made by this framework is between predictions and postdictions, that is, the time when learners make metacomprehension judgments. Predictions are made after learners have read a text but before they take a comprehension test. In contrast, postdictions are made after learners have taken a comprehension test. According to the framework, when making predictions, learners can use a variety of cues for judging their text comprehension. These cues, however, differ in the extent to which they are valid indicators of text comprehension. For example, learners can base their judgments on heuristic cues such as topic interest or domain familiarity. As these cues are available regardless whether a text has been read or not, they do not closely relate to the construction of a particular mental text representation and, thus, are usually no valid indicators of text comprehension. Therefore, when learners take a heuristic route to judgment, metacomprehension might be rather inaccurate. Alternatively, learners can base their judgments on representation-based cues such as accessibility or memorability of textual information. As these cues closely relate to the construction of a particular mental text representation, they are typically valid indicators of text comprehension. Thus, when learners take a representation-based route to judgment, metacomprehension should be fairly accurate.
The representation-based route to judgment might nevertheless be misleading when learners possess misconceptions. More concretely, Dunlosky et al. (2005) showed that learners often focus on the amount of textual information they can access from memory as a basis for making predictions but, at the same time, neglect the quality of the accessed information. If learners have misconceptions, however, a large proportion of textual information they access from memory will likely be incorrect (Kendeou and van den Broek 2005). Hence, when these learners focus on the amount of textual information accessed as a judgment cue, they make predictions that do not adequately mirror their text comprehension. More precisely, in this case the learners are likely to overestimate their comprehension. Indirect evidence for this assumption comes from two experiments conducted by van Loon et al. (2013) in which children were given the task to learn new concepts such as fossil. The results showed that children with inaccurate prior knowledge overestimated their understanding because they made fairly optimistic predictions, whereas their actual understanding of the concepts was rather limited. Obviously, when making predictions, children with inaccurate prior knowledge retrieved information from memory but failed to sufficiently pay attention to the quality of this information (Van Loon et al. 2013). However, it is important to mention that the reported studies on the impact of prior knowledge on monitoring accuracy (Griffin et al. 2009; Schneider et al. 1989; Van Loon et al. 2013) focused on conceptual learning contents. In general, research on monitoring accuracy with regard to procedural contents in educational settings is scarce (see Baars et al. 2013, 2014a, 2014b; De Bruin et al. 2005, 2007 for exceptions) and completely missing in the context of learning from text.

Furthermore, according to the cue-utilization framework (Griffin et al. 2009), the postdiction route to judgment deals with metacomprehension judgments made after learners have completed a comprehension test. In this case, learners might use cues that are available
from the heuristic and representation-based route to a lesser extent. Instead, they might primarily rely on cues that directly come from completing comprehension questions such as the frequency of guessing, perceptions of the plausibility or difficulty of the distractors of the questions, (dis)fluency during completing the questions, effort invested in completing the questions, and feelings of failure or success in completing the questions (cf. Koriat et al. 2006). As metacomprehension accuracy is defined as judgment–performance correspondence, cues that are proximally tied to comprehension performance yield most accurate judgments (Griffin et al. 2009). Thus, as supported by the postdiction superiority effect (Pierce and Smith 2001), postdictions should be more accurate than predictions. Hence, in case of postdictions, misconceptions might not play a remarkable role for the accuracy of learners’ judgments any longer. Through completing comprehension questions, various cues that should be valid indicators of comprehension become available to the learners and might undermine the detrimental impact of misconceptions on how they judge their comprehension. As a result, learners with misconceptions might become more accurate in their postdictions than in their predictions (cf. Maki et al. 2005).

**The Present Study**

In this study, we examined whether misconceptions would adversely affect text comprehension and metacomprehension accuracy with regard to predictions and postdictions in the domain of statistics. In doing so, we focused on the topic of covariance. As learning in statistics usually includes the acquisition of both concepts and procedures, text comprehension and metacomprehension accuracy referred to both conceptual and procedural aspects of covariance.

First, we investigated to what extent misconceptions interfered with the comprehension of a statistics text. We predicted that learners with a higher number of misconceptions would show poorer conceptual and procedural text comprehension than
learners with a lower number of misconceptions (Hypothesis 1a). Conceptual and procedural knowledge are usually highly interrelated. Therefore, we expected the negative influence of misconceptions to hold true for both conceptual and procedural text comprehension. However, we assumed that the detrimental effect of misconceptions on text comprehension would be particularly pronounced with regard to conceptual comprehension because this type of comprehension is more closely related to misconceptions about covariance than procedural comprehension (Hypothesis 1b).

Second, we investigated to what extent misconceptions impaired prediction accuracy. We assumed that learners with a higher number of misconceptions would show greater overestimation of their conceptual and procedural text comprehension than learners with a lower number of misconceptions (Hypothesis 2a). Again, we assumed that misconceptions would have a larger effect on prediction accuracy of conceptual comprehension than on prediction accuracy of procedural comprehension (Hypothesis 2b).

Third, we investigated to what extent misconceptions affected postdiction accuracy. In contrast to predictions, postdictions are mainly influenced by the information available from completing comprehension questions (e.g., frequency of guessing). Therefore, we expected that learners’ misconceptions would have a smaller (or even no) effect on postdiction accuracy than on prediction accuracy with regard to both conceptual and procedural text comprehension (Hypothesis 3a). In addition, we assumed the difference in the impact of misconceptions on postdiction accuracy of conceptual comprehension and postdiction accuracy of procedural comprehension to be rather small (or even zero) because misconceptions should not play an important role for the postdictions of both comprehension types (Hypothesis 3b).
Method

Sample

A total of $N = 47$ university students enrolled in educational science, cognitive science, or psychology participated in this study. The mean age of the participants was 24.04 ($SD = 5.43$) years with the vast majority (83%) being female. To ensure that participants had some prior knowledge and, thus, could possess misconceptions about covariance, only students who had attended at least one statistics course that covered the topic of covariance were eligible to take part in this study. On average, the participants had attended 2.75 ($SD = 0.90$) statistics courses. The participants received either course credit or pay for taking part in the study.

Design

The study had a correlational design with number of misconceptions about covariance as the predictor variable. The dependent variables were text comprehension and metacomprehension accuracy with regard to predictions and postdictions. Text comprehension and metacomprehension accuracy referred to both conceptual and procedural aspects of covariance.

Materials and Measures

Statistics text. The statistics text that participants read in this study was adapted from a statistics textbook written by Bortz and Schuster (2010) and covered the topic of covariance as a measure of association (see Appendix A for the text). The text addressed conceptual aspects of covariance, for instance, its different directions, as well as procedural aspects of covariance, for example, how it is calculated. Moreover, the text contained three graphs to illustrate positive, negative, and no covariance as well as the formula to calculate covariance. In total, the text comprised 564 words. Without the graphs and the formula, the text had a
Flesch-Reading-Ease score (Flesch 1948) of 33, which indicates that the text was rather difficult to read.

**Misconceptions.** Participants’ misconceptions about covariance were assessed via 15 questions (see Appendix B for the misconceptions test). Each of the questions addressed one particular misconception. As a basis for constructing these questions, we reviewed the literature on preconceptions and misconceptions about measures of association (e.g., Batanero et al. 1997; Estepa and Batanero 1996; Liu et al. 2009; Morris 1999). This search yielded 15 misconceptions that individuals might hold about the concept of covariance (see Appendix C for the misconceptions with the respective references). Each of the 15 questions had a single-choice format with four response options. One option represented the correct answer, one option represented the particular misconception, and the two remaining options represented incorrect answers but not a misconception. The number of misconceptions was determined by counting how many times a participant selected the response option that represented a misconception. Thus, a participant could have a maximum number of 15 misconceptions. The test consisted of rather distinct misconceptions. Therefore, participants could have one misconception without necessarily having another misconception. Thus, not unexpectedly, internal consistency for the misconceptions test was rather low, $\alpha = .31$ (cf. Van Loon et al. 2015).

**Text comprehension.** Text comprehension referred to both conceptual and procedural aspects of covariance presented in the statistics text. Conceptual comprehension was assessed by eight inference questions the correct answers to which were not explicitly stated in the text but had to be deduced by the participants (see Appendix D for the conceptual comprehension questions). Of these eight questions, four questions were taken from the misconceptions test and, thus, addressed misconceptions about covariance as described above. The remaining four questions addressed further conceptual attributes of covariance but not specifically
misconceptions. These four questions also had a single-choice format with four response options. One option represented the correct answer and three options represented incorrect answers. The participants received 1 point for a correct answer on the conceptual comprehension questions. Thus, they could achieve a maximum number of 8 points. Internal consistency for the conceptual comprehension questions was rather low, $\alpha = .40$. The rather low alpha coefficient reflects the fact that the conceptual comprehension questions referred to multiple unrelated conceptual aspects of covariance. Participants could know one aspect without necessarily knowing another aspect. We constructed the questions this way because we wanted to assess comprehension of covariance in a rather comprehensive manner. Therefore, the obtained alpha coefficient is acceptable in the context of this study (cf. Schneider et al. 2011). Moreover, the fact that conceptual comprehension significantly correlated with reading skills (see Table 2) indicates that the questions indeed measured conceptual comprehension of the statistics text about covariance.

Procedural comprehension was assessed by four questions that required calculations regarding covariance (see Appendix E for the procedural comprehension questions). All four procedural comprehension questions had an open-ended format and asked for the solution steps and the final solution. The participants received 1 point for a correct answer. Thus, they could achieve a maximum number of 4 points. Two raters independently scored participants’ answers to the procedural comprehension questions. Interrater agreement was high, Cohen’s $\kappa = .97$, 95% CI [0.93, 1.00] (Cohen 1960). Internal consistency for the procedural comprehension questions was good, $\alpha = .73$.

To facilitate the interpretation of participants’ performance on the conceptual and procedural comprehension questions, which were unequal in count, we converted the number of conceptual and procedural comprehension questions correct into percent correct. There
was one missing value for the conceptual and procedural comprehension questions, respectively. These missing values were treated as errors and, hence, assigned 0 points.

**Metacomprehension accuracy.** We assessed metacomprehension accuracy for both participants’ predictions and postdictions of their text comprehension. To predict their text comprehension, the participants received the following instructions:

You will receive a total of 12 questions about the text you have just read.

The first 8 questions are conceptual questions that address your understanding of various conceptual aspects of covariance. For each question, there are four response options of which only one is correct. Please tick the response option that you consider to be the correct answer. You will have 12 minutes to complete these questions.

The next 4 questions are procedural questions that require you to perform calculations with regard to covariance. Please answer each question by recording both the solution steps and the final solution. Again, you will have 12 minutes to complete these questions.

Before you complete the questions, please estimate how many of the questions you will presumably be able to answer correctly.

**Your judgment:**

Of the 8 conceptual questions, I will answer _____ questions correctly.

Of the 4 procedural questions, I will answer _____ questions correctly.

To postdict their text comprehension, the participants judged the number of questions they presumably answered correctly after they had completed the questions. Again, they did
so for the conceptual comprehension questions (“Of the 8 conceptual questions, I answered _____ questions correctly.”) and for the procedural comprehension questions (“Of the 4 procedural questions, I answered _____ questions correctly.”) separately. There was one missing value for the postdiction of procedural comprehension that was excluded from the statistical analyses.

To test our hypotheses, we operationalized metacomprehension accuracy with regard to predictions and postdictions in terms of bias. Bias referred to the signed difference between a participant’s judged number of questions correct (in %) and actual number of questions correct (in %; see, e.g., Schraw 2009). A positive value indicated overconfidence, that is, a participant provided a higher judgment compared to his or her actual performance. For example, a value of .25 meant that a participant assumed to provide 25% more correct answers than it was actually the case. In contrast, a negative value indicated underconfidence, that is, a participant provided a lower judgment compared to his or her actual performance. A value of zero indicated a perfectly accurate judgment.

However, a limitation of bias is that its average cannot be used to investigate the mean accuracy of the sample. When average bias is computed for the sample and there are participants who are overconfident and participants who are underconfident, over- and underconfidence cancel each other out. Therefore, to investigate mean accuracy, we additionally operationalized metacomprehension accuracy in terms of absolute accuracy (cf., e.g., Baars et al. 2013, 2014a, 2014b; Maki et al. 2005). Absolute accuracy referred to the absolute difference between a participant’s judged number of questions correct (in %) and actual number of questions correct (in %), regardless of the direction of the difference (see, e.g., Schraw 2009). Thus, for absolute accuracy, there were only positive values, with a higher value indicating greater inaccuracy. A value of zero indicated a perfectly accurate judgment.
**Reading skills.** Prior research has shown that reading skills play an important role for text comprehension and metacomprehension (e.g., Kwon and Linderholm 2014; Perfetti 1985). Therefore, we also assessed participants’ reading skills with a subtest of the computer-based ELVES test (Efficiency of Reading Comprehension in Adult Readers According to the Strategy Model; Richter and van Holt 2005). In this subtest, the participants had to judge the meaningfulness of sentences that varied in syntactic and semantic complexity (e.g., “If one eats spoiled food, the skin often discolors sonorously.”) as quickly as possible. This subtest assesses lower-order reading processes such as extracting propositions and integrating them syntactically and semantically on the sentence level. We chose this subtest because it consumes relatively little administration time and was found to be strongly correlated with higher-order reading processes such as identifying text implications (Richter and van Holt 2005). Therefore, the subtest should provide a valid indication of participants’ reading skills.

**Procedure**

The study was divided into three phases. In the first phase, the participants completed the misconceptions test within a time limit of 20 minutes. Afterwards, they accomplished the reading skills test, which also served as a filler task to remove the content of the misconceptions test from working memory. In the second phase, the participants read the statistics text about covariance. They were instructed to carefully read the text in order to understand the information provided. In addition, they were informed that their conceptual and procedural comprehension of the text would be tested after reading. While reading, it was not allowed to highlight text or take notes. The time for reading was limited to 8 minutes. In the third phase, the statistics text was removed and the participants made the predictions of their conceptual and procedural text comprehension. To help the participants understand what was meant by conceptual and procedural comprehension questions, they were informed about the kind of knowledge that the two types of comprehension questions would require. Then,
the participants answered the conceptual comprehension questions within a time limit of 12 minutes and, afterwards, made the postdiction of their conceptual text comprehension. Likewise, the participants answered the procedural comprehension questions within a time limit of 12 minutes and, afterwards, made the postdiction of their procedural text comprehension. Finally, they answered some demographic questions.

**Results**

To statistically test our hypotheses, we performed regression analyses in which we controlled for the influence of reading skills. More precisely, we subjected the dependent variables to linear regression analyses with number of misconceptions about covariance as the predictor and reading skills as the covariate. We controlled for the influence of reading skills as it appeared to be significantly correlated with text comprehension and postdictions in this study (see Table 2). For all statistical analyses, we used an alpha level of .05. When testing directional hypotheses, we used one-tailed tests.

**Descriptive Statistics for Misconceptions and the Dependent Variables**

The average number of misconceptions that participants held was 3.96 misconceptions ($SD = 1.93$; see Table 1 for an overview of the number of participants who held each of the 15 misconceptions). The relatively high standard deviation shows that there was quite a large variability among the participants. In fact, one participant had no misconceptions at all and one participant had a maximum number of nine misconceptions. Misconceptions were not correlated with reading skills ($r = -.16, p = .274$). This indicates that participants who had more misconceptions were not simultaneously the ones who had poorer reading skills.

--- Please insert Table 1 about here ---

Table 2 provides the intercorrelations and descriptive statistics for the dependent variables. Average bias cannot be meaningfully interpreted because overconfident judgments
and underconfident judgments cancel each other out. Thus, to evaluate mean accuracy, we used absolute accuracy instead of bias. As shown in Table 2, concerning the absolute accuracy of predictions, participants misjudged their conceptual comprehension on average by 14% (SD = 13%) and their procedural comprehension on average by 31% (SD = 22%). Moreover, concerning the absolute accuracy of postdictions, participants misjudged their conceptual comprehension on average by 15% (SD = 11%) and their procedural comprehension on average by 18% (SD = 17%).

--- Please insert Table 2 about here ---

**Misconceptions and Text Comprehension**

Hypothesis 1a predicted that misconceptions would negatively affect conceptual and procedural text comprehension. As depicted in Table 3, the linear regression analyses on text comprehension showed that misconceptions in fact influenced conceptual comprehension significantly. More concretely, an increase of one misconception resulted in a decrease of 3% in comprehension. In contrast, misconceptions failed to be a significant predictor of procedural comprehension. Thus, Hypothesis 1a was confirmed with regard to conceptual text comprehension but not with regard to procedural text comprehension.

--- Please insert Table 3 about here ---

Hypothesis 1b assumed that the negative impact of misconceptions would be more pronounced for conceptual comprehension than for procedural comprehension. However, there was no significant difference between the influence of misconceptions on conceptual comprehension and the influence of misconceptions on procedural comprehension, $b = 0.01$, $SE = 0.02$, $\beta = 0.03$, $t(44) = 0.23$, $p = .411$. Thus, the results were not in line with Hypothesis 1b.
**Misconceptions and Judgment Magnitudes**

We first analyzed the effects of misconceptions on the magnitudes of participants’ predictions and postdictions per se before we examined accuracy. As shown in Table 4, the linear regression analyses on predictions revealed that misconceptions had no significant influence on the prediction of either conceptual comprehension or procedural comprehension.

--- Please insert Table 4 about here ---

As represented in Table 5, the linear regression analyses on postdictions showed that misconceptions significantly influenced the postdiction of conceptual comprehension with an increase of one misconception leading to a decrease of 3% in judgment magnitude. In contrast, misconceptions did not significantly affect the postdiction of procedural comprehension.

--- Please insert Table 5 about here ---

In sum, misconceptions did not significantly influence participants’ predictions. This result indicates that participants with misconceptions were not sensitive to their poor text comprehension when providing their predictions. In contrast, the more misconceptions participants had, the lower they postdicted their conceptual text comprehension to be. This, however, was not true for the postdiction of their procedural text comprehension.

**Misconceptions and Prediction Accuracy**

Concerning the prediction of conceptual text comprehension, bias scores ranged from -38% (i.e., underestimation) to +38% (i.e., overestimation). Concerning the prediction of procedural text comprehension, bias scores ranged from -50% (i.e., underestimation) to +100% (i.e., overestimation). Thus, there was a large variability in the prediction bias scores.

Hypothesis 2a supposed that more misconceptions would result in greater overestimation when participants predicted their conceptual and procedural text comprehension. As our hypothesis concerned the influence of misconceptions on the
direction of inaccuracy in learners’ predictions (i.e., under-/overestimation), we selected bias as the accuracy measure for the analyses. However, the analyses with absolute accuracy are provided in Appendix F for completeness. As shown in Table 6, the linear regression analyses on prediction biases revealed that misconceptions significantly influenced both the accuracy with which participants judged their conceptual comprehension and the accuracy with which they judged their procedural comprehension. More concretely, an increase of one misconception resulted in 3% less underestimation or greater overestimation of conceptual comprehension and 6% less underestimation or greater overestimation of procedural comprehension. The patterns of these relationships are depicted in Figure 1. To conclude, Hypothesis 2a was confirmed because the higher the participants’ number of misconceptions was, the more they overestimated their text comprehension. Importantly, this effect occurred when participants predicted their conceptual as well as their procedural comprehension.

--- Please insert Table 6 about here ---

--- Please insert Figure 1 about here ---

**Fig. 1** Effects of number of misconceptions on prediction biases controlled for reading skills

According to Hypothesis 2b, overconfidence due to misconceptions should be more pronounced for the prediction of conceptual comprehension than for the prediction of procedural comprehension. However, Hypothesis 2b was not confirmed because there was no significant difference between the influence of misconceptions on prediction bias of conceptual comprehension and the influence of misconceptions on prediction bias of procedural comprehension, \( b = -0.03, SE = 0.03, \beta = -.15, t(44) = -1.02, p = .158. \)

**Misconceptions and Postdiction Accuracy**

Concerning the postdiction of conceptual text comprehension, bias scores ranged from -50% (i.e., underestimation) to +38% (i.e., overestimation). Concerning the postdiction of procedural text comprehension, bias scores ranged from -50% (i.e., underestimation) to +50%
(i.e., overestimation). Thus, there was a rather high variability in the postdiction bias scores as well.

Hypothesis 3a assumed that postdiction accuracy of both conceptual and procedural text comprehension would be less affected by misconceptions than prediction accuracy. Again, we selected bias as the accuracy measure for the analyses (for the analyses with absolute accuracy, see Appendix F). As depicted in Table 7, the linear regression analyses on postdiction biases revealed that misconceptions neither significantly influenced the accuracy with which participants judged their conceptual comprehension nor the accuracy with which they judged their procedural comprehension (see also Figure 2). Hence, in line with Hypothesis 3a, postdiction accuracy with regard to both conceptual and procedural text comprehension was not affected by misconceptions.

--- Please insert Table 7 about here ---

--- Please insert Figure 2 about here ---

Fig. 2 Effects of number of misconceptions on postdiction biases controlled for reading skills

According to Hypothesis 3b, misconceptions should affect postdiction accuracy of conceptual comprehension and postdiction accuracy of procedural comprehension to the same (small) degree. In line with this hypothesis, there was no significant difference between the influence of misconceptions on postdiction bias of conceptual comprehension and the influence of misconceptions on postdiction bias of procedural comprehension, $b = -0.02, SE = 0.02, \beta = -0.12, t(43) = -0.76, p = .454$.

We also investigated whether the accuracy with which participants judged their text comprehension changed from predictions to postdictions depending on the number of misconceptions. To do so, we subtracted prediction bias from postdiction bias and regressed this difference on the number of misconceptions while controlling for reading skills. As the comprehension variable used to calculate bias was the same for prediction bias and
postdiction bias, a larger difference between prediction bias and postdiction bias indicated a greater change in judgment magnitude. A positive difference indicated that a participant had increased his or her judgment from prediction to postdiction, whereas a negative difference indicated that that a participant had decreased his or her judgment from prediction to postdiction. Concerning conceptual comprehension, the more misconceptions participants had, the greater was the difference between prediction bias and postdiction bias in negative direction, \( b = -0.03, SE = 0.01, \beta = -0.32, t(44) = -2.27, p = .028 \). In other words, participants with a higher number of misconceptions adapted their judgments downwards in their postdictions of conceptual comprehension compared with their predictions of conceptual comprehension, which means that they became less overconfident. In contrast, participants with a lower number of misconceptions rather adapted their judgments upwards in their postdictions of conceptual comprehension compared with their predictions of conceptual comprehension, which means that they became less underconfident (also refer to Figure 1 and Figure 2 to see the change in accuracy from predictions to postdictions). Concerning procedural comprehension, the effect of misconceptions on the difference between prediction bias and postdiction bias was statistically not significant, \( b = -0.03, SE = 0.03, \beta = -0.20, t(43) = -1.36, p = .180 \).

**Discussion**

In line with previous research (e.g., Liu et al. 2009; Morris 1999; see also Huck 2016), we found that the learners in this study had a quite large number of statistical misconceptions, for example, that covariance proves causality. We investigated whether these misconceptions impaired not only their comprehension of a text on statistics but also the accuracy with which they judged their comprehension of this text.

First, our results showed that a higher number of misconceptions led to poorer comprehension of the conceptual aspects of the statistics text. This is in line with prior
research demonstrating that misconceptions hamper the comprehension of expository text (e.g., Kendeou and van den Broek 2005). Yet, prior research has mainly focused on the impact of misconceptions on comprehending science texts (see, e.g., Guzzetti et al. 1993; Tippett 2010). Therefore, our finding extends prior research by showing that also in the domain of statistics misconceptions are detrimental to text comprehension. However, contrary to expectation, misconceptions did not significantly influence the comprehension of procedural aspects of the statistics text. Even though procedural comprehension was significantly associated with conceptual comprehension in this study (see Table 2), procedural comprehension obviously referred to statistical aspects that were not immediately affected by statistical misconceptions. An explanation for this finding is that misconceptions about covariance primarily address conceptual aspects. Therefore, they might be more intimately related to conceptual comprehension than to procedural comprehension.

Second, participants’ misconceptions affected the accuracy of their predictions of conceptual and procedural text comprehension. This finding extends prior research by demonstrating that not only the amount of prior knowledge (Griffin et al. 2009; Schneider et al. 1989) but also the correctness of prior knowledge in terms of misconceptions influences the accuracy with which learners predict their text comprehension.

Why participants with more misconceptions overestimated their text comprehension more strongly than participants with fewer misconceptions can be explained with the help of the cue-utilization framework (Griffin et al. 2009; cf. Koriat 1997). It is plausible to assume that all participants in our study constructed a mental representation of the information provided in the statistics text. However, if participants had a larger number of misconceptions, their mental text representation was flawed. Hence, when these participants took a representation-based route to judgment but based their predictions mainly on the amount of textual information they could retrieve from memory, they were likely to
overestimate their actual comprehension. Yet, to shed more light on the mechanisms underlying this overestimation bias, future research is encouraged to examine in more detail the processes involved in forming metacomprehension judgments by learners who have misconceptions. The finding that learners who had more misconceptions were more strongly overconfident suggests that these learners run the risk of abandoning further learning activities, such as rereading a text, that would enhance their text comprehension.

That participants with fewer misconceptions tended to underestimate their text comprehension might reflect general low self-efficacy when it comes to learning statistics (e.g., Perepiczka et al. 2011). Therefore, although participants with a lower number of misconceptions might have retrieved a large amount of correct textual information from memory when judging their comprehension, they might have provided restrained predictions compared with their actual comprehension.

The finding that misconceptions affected prediction accuracy of conceptual and procedural comprehension stands in contrast to the finding that misconceptions influenced only conceptual comprehension but not procedural comprehension. An explanation for why misconceptions impaired accuracy when participants predicted both types of comprehension is that learners are hardly sensitive to varying conditions, such as text difficulty, when self-assessing their comprehension (e.g., Maki et al. 2005). Accordingly, in this study, participants might have produced very similar judgments when assessing their conceptual and procedural text comprehension. This assumption is corroborated by the statistically strong association between the prediction of conceptual comprehension and the prediction of procedural comprehension (see Table 2).

Third, this study revealed that misconceptions exerted no influence on the accuracy with which participants postdicted their conceptual and procedural text comprehension. Misconceptions played no role for postdiction accuracy, which supports the notion that, in
contrast to predictions, postdictions do not primarily mirror monitoring processes related to comprehending a text itself but, at least partly, are the result of monitoring processes associated with answering comprehension questions (Griffin et al. 2009). More precisely, for their postdictions, participants could use cues that resulted from completing the comprehension questions such as the frequency of guessing, perceptions of how plausible or difficult the distractors of the questions were, (dis)fluency during completing the questions, effort invested in completing the questions, and feelings of failure or success in completing the questions (cf. Koriat et al. 2006). In particular participants with misconceptions seem to have benefited from such cues so that they were less prone to focus on the amount of textual information retrieved from memory as a judgment cue. This assumption is supported by the finding that participants with a higher number of misconceptions adapted their judgments downwards in their postdictions of conceptual comprehension compared with their respective predictions so that they were not more overconfident than participants with a lower number of misconceptions any longer. This finding suggests that learners with misconceptions are not sensitive to flaws in comprehension after reading a text. However, after having completed comprehension questions, they obviously become aware that their understanding is limited. Thus, practice tests seem to be an effective instructional method to promote learners with misconceptions in accurately judging their text comprehension that would enable them to effectively self-regulate their learning. Our results suggest that through completing the comprehension questions the participants had access to cues that entailed a higher degree of validity; however, it remains open which specific cues they used. Consequently, it is unclear to what extent the benefit of answering comprehension questions in fact depends on altered cue utilization. To address this issue, future studies should assess learners’ cue utilization by using, for example, standardized interviews or the think-aloud methodology.
Limitations and Future Directions

The present research contributes to an advanced understanding about the role of misconceptions in learning from text. Our study suggests that learners with misconceptions indeed suffer from a double curse: Their misconceptions not only impair their comprehension of a text but also prevent them from accurately self-assessing their comprehension. However, it is important to highlight some limitations of this study. First, metacomprehension accuracy was measured in terms of bias and absolute accuracy. However, these two measures, and bias in particular, statistically rely on the level of performance. Therefore, not only the ability to judge comprehension but also the level of performance in a comprehension test can influence metacomprehension accuracy (Griffin et al. 2009; for a longer discussion, see Nelson 1984). For example, when a learner achieves a rather high level of performance in a comprehension test, the range for making underconfident judgments is much larger than the range for making overconfident judgments. Relative accuracy is another measure of metacomprehension accuracy and refers to the extent to which metacomprehension judgments correlate with actual comprehension performance across several texts (see, e.g., Schraw 2009). Therefore, it is less affected by the performance level than bias and absolute accuracy (Griffin et al. 2009). Consequently, future research might use relative accuracy as an additional measure to examine the role of misconceptions for metacomprehension accuracy. It is noteworthy, however, that relative accuracy has some methodological shortcomings, too. For example, it can be constrained by low within-person variance in performance or metacomprehension judgments. In addition, relative accuracy assesses other aspects of metacomprehension accuracy than bias and absolute accuracy. More specifically, relative accuracy provides information about the extent to which learners can accurately judge their comprehension on one text relative to other texts, whereas absolute accuracy provides information concerning the precision of learners’ judgments of comprehension, and bias provides unique information.
about whether learners are over- or underconfident when judging their comprehension (Schraw 2009). This information is highly important given that overconfidence can lead to the premature termination of study and, thus, may have implications for the success of learning (e.g., Dunlosky and Rawson 2012). Moreover, it has been recommended to use bias as the measure of accuracy to investigate whether judgments decrease or increase relative to performance (Schraw 2009).

Second, as already mentioned, the mechanisms underlying the effects observed in this study remain rather unclear. Therefore, future research should combine offline measures, such as metacomprehension judgments, with online measures, such as think-aloud protocols, eye-tracking techniques, or reading and processing times, to provide further insights into how learners with varying degrees of misconceptions read texts and monitor their text comprehension (cf., e.g., Kendeou and van den Broek 2005, 2007). For example, having participants thinking aloud during reading, test taking, and making judgments, could unveil the cues they use to make predictions and postdictions. Besides, attribution interviews (cf. Bol et al. 2005) could be useful to investigate the cues learners attribute their (mis)judgments to and how they relate to their misconceptions.

Third, the great majority of the participants in this study were female. Therefore, the generalizability of the results to populations that include more males should be further examined. For example, prior research suggests that men hold fewer misconceptions in statistics than women (Liu and Garfield 2002).

Fourth and finally, the results of this study are based on correlational data. Therefore, causal conclusions concerning the impact of misconceptions on text comprehension and metacomprehension accuracy should be treated with caution. However, some aspects of this study suggest that misconceptions indeed caused poorer text comprehension and metacomprehension accuracy. One aspect is that the misconceptions test preceded the
metacomprehension judgments and comprehension questions. Thus, the temporal sequence of cause (i.e., misconceptions) and effect (i.e., text comprehension, metacomprehension accuracy) is given, supporting the assumed directionality of the observed effects (Kenny 2004). Moreover, by controlling for reading skills, we ruled out the possibility that the observed variation in comprehension and metacomprehension accuracy attributed to differences in the number of misconceptions was associated with this important variable in the context of learning from text (e.g., Kwon and Linderholm 2014; Perfetti 1985).

**Improving Text Comprehension and Metacomprehension Accuracy**

The results obtained in this study call for methods to support learners with misconceptions in learning from text. A promising way to remove misconceptions and foster text comprehension is the use of refutation texts instead of conventional expository texts (e.g., Kendeou and van den Broek 2007; see also Guzzetti et al. 1993; Tippett 2010). Refutation texts explicitly state commonly held misconceptions about a topic and directly refute them with the scientifically correct explanations (Tippett 2010). However, it remains open whether refutation texts are beneficial for learning in the domain of statistics.

In addition, to improve the accuracy with which learners self-assess their text comprehension, a variety of techniques can come into play. For example, generating keywords and summarizing text have been shown to increase metacomprehension accuracy (see, e.g., Thiede et al. 2009). In case learners have misconceptions, however, it seems to be particularly important to help them to become aware of their flawed knowledge. Otherwise, they are likely to produce overconfident predictions. Therefore, by supporting learners in reflecting about their misconceptions, refutation texts might also prove beneficial for metacomprehension accuracy. Whether this type of text improves not only comprehension but also metacomprehension accuracy should be examined by future research.
References


Erfassung der Effizienz von Teilprozessen des Leseverstehens [ELVES: Computer-based 
assessment of the efficiency of component processes involved in reading comprehension]. 

knowledge of mathematics. In R. C. Kadosh & A. Dowker (Eds.), *The Oxford handbook 

knowledge, procedural knowledge, and procedural flexibility in two samples differing in 
prior knowledge. *Developmental Psychology, 47*(6), 1525–1538. doi:10.1037/a0024997

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory 

*Metacognition and Learning, 4*(1), 33–45. doi:10.1007/s11409-008-9031-3

Sundre, D. L. (2003, April). Assessment of quantitative reasoning to enhance educational 
quality. Paper presentation at the annual American Educational Research Association, 
Chicago, IL. Retrieved from 

monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. 
doi:10.1037/0022-0663.95.1.66

during and after reading. In A. C. Graesser, D. J. Hacker, & J. Dunlosky (Eds.), *Handbook 


### Table 1

**Absolute and Relative Frequency of the 15 Misconceptions**

<table>
<thead>
<tr>
<th>Misconception</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1. Covariance enables definite predictions</td>
<td>17</td>
<td>36</td>
</tr>
<tr>
<td>M2. Negative covariance indicates the absence of a relationship</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>M3. Covariance is related to the slope of the fit line</td>
<td>20</td>
<td>43</td>
</tr>
<tr>
<td>M4. A straight fit line without deviation invariably indicates perfect covariance</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M5. Covariance implies causality</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>M6. Positive covariance is always stronger than negative covariance</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td>M7. Covariance can be interpreted from isolated points in scatter plots</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>M8. Zero covariance indicates the absence of any association</td>
<td>29</td>
<td>62</td>
</tr>
<tr>
<td>M9. Only consider covariance if attributable to a causal relationship</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>M10. Only a straight fit line without deviation indicates covariance</td>
<td>13</td>
<td>28</td>
</tr>
<tr>
<td>M11. Positive and zero covariance are always stronger than negative covariance</td>
<td>13</td>
<td>28</td>
</tr>
<tr>
<td>M12. Covariance is a standardized statistic</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td>M13. Covariance changes if the variables are reversed</td>
<td>7</td>
<td>15</td>
</tr>
</tbody>
</table>
M14. More negative covariance indicates a weaker relationship

M15. The fit line of perfect covariance must have a constant slope of 45 degrees
Table 2

*Intercorrelations, Means, and Standard Deviations for Text Comprehension, Judgment Magnitudes, Accuracy Measures, Misconceptions, and Reading Skills*

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.72</td>
<td>.17</td>
</tr>
<tr>
<td>2. PC</td>
<td>.44*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.63</td>
<td>.35</td>
</tr>
<tr>
<td>3. Prediction CC</td>
<td>.34*</td>
<td>.40*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.71</td>
<td>.15</td>
</tr>
<tr>
<td>4. Prediction PC</td>
<td>.19</td>
<td>.13</td>
<td>.63*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.61</td>
<td>.21</td>
</tr>
<tr>
<td>5. Postdiction CC</td>
<td>.42*</td>
<td>.25</td>
<td>.64*</td>
<td>.35*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.69</td>
<td>.17</td>
</tr>
<tr>
<td>6. Postdiction PC</td>
<td>.43*</td>
<td>.75*</td>
<td>.53*</td>
<td>.30*</td>
<td>.39*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.70</td>
<td>.31</td>
</tr>
<tr>
<td>7. Bias prediction CC</td>
<td>-.66*</td>
<td>-.09</td>
<td>.49*</td>
<td>.33*</td>
<td>.12</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.01</td>
<td>.19</td>
</tr>
<tr>
<td>8. Bias prediction PC</td>
<td>-.30*</td>
<td>-.84*</td>
<td>-.01</td>
<td>.44*</td>
<td>-.04</td>
<td>-.51*</td>
<td>.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.02</td>
<td>.39</td>
</tr>
<tr>
<td>9. Bias postdiction CC</td>
<td>-.54*</td>
<td>-.18</td>
<td>.28</td>
<td>.15</td>
<td>.54*</td>
<td>-.04</td>
<td>.72*</td>
<td>.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.03</td>
<td>.19</td>
</tr>
<tr>
<td>10. Bias postdiction PC</td>
<td>-.10</td>
<td>-.50*</td>
<td>.09</td>
<td>.19</td>
<td>.14</td>
<td>.20</td>
<td>.16</td>
<td>.56*</td>
<td>.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.07</td>
<td>.24</td>
</tr>
<tr>
<td>11. Absolute accuracy</td>
<td>.05</td>
<td>.08</td>
<td>.03</td>
<td>.01</td>
<td>.03</td>
<td>-.11</td>
<td>-.02</td>
<td>-.07</td>
<td>-.02</td>
<td>-.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.14</td>
</tr>
</tbody>
</table>
12. Absolute accuracy prediction PC
|       | .06 | -.07 | .11 | -.08 | -.11 | .04 | .18 | -.13 | -.04 | .10 | – | .31 | .22 |

13. Absolute accuracy postdiction CC
|       | .15 | .09 | .01 | -.03 | -.17 | -.14 | -.13 | -.10 | -.29* | -.32* | .37* | .27 | – | .15 | .11 |

14. Absolute accuracy postdiction PC
|       | .02 | -.29 | .09 | .15 | .09 | -.02 | .06 | .34* | .06 | .39* | -.12 | -.17 | -.16 | – | .18 | .17 |

15. Number of misconceptions
|       | -.41* | -.26 | -.14 | .12 | -.40* | -.16 | .27 | .30* | .01 | .17 | -.01 | .07 | .09 | .13 | – | 3.96 | 1.93 |

16. Reading skills
|       | .30* | .32* | .25 | .17 | .39* | .29* | -.08 | -.20 | .09 | -.09 | -.02 | -.21 | -.17 | .04 | -.16 | – | 17.08 | 6.32 |

*Note. CC = conceptual comprehension; PC = procedural comprehension.

*p < .05.
Table 3

*Regression Analyses for Text Comprehension*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>b</th>
<th>SE</th>
<th>β</th>
<th>t(44)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conceptual Comprehension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.74</td>
<td>0.09</td>
<td>8.42</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>Reading skills</td>
<td>0.01</td>
<td>0.01</td>
<td>0.24</td>
<td>1.74</td>
<td>.089</td>
</tr>
<tr>
<td>Number of misconceptions</td>
<td>-0.03</td>
<td>0.01</td>
<td>-0.37</td>
<td>-2.71</td>
<td>.005</td>
</tr>
<tr>
<td><strong>Procedural Comprehension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.51</td>
<td>0.19</td>
<td>2.73</td>
<td>.009</td>
<td></td>
</tr>
<tr>
<td>Reading skills</td>
<td>0.02</td>
<td>0.01</td>
<td>0.29</td>
<td>2.02</td>
<td>.049</td>
</tr>
<tr>
<td>Number of misconceptions</td>
<td>-0.04</td>
<td>0.03</td>
<td>-0.21</td>
<td>-1.49</td>
<td>.072</td>
</tr>
</tbody>
</table>

\[ R^2 = .22, F(2, 44) = 6.12, p = .005 \]

\[ R^2 = .15, F(2, 44) = 3.76, p = .031 \]
Table 4

*Regression Analyses for Prediction Magnitudes*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$b$</th>
<th>SE</th>
<th>$\beta$</th>
<th>$t(44)$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prediction of Conceptual Comprehension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.64</td>
<td>0.08</td>
<td>7.80</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>Reading skills</td>
<td>0.01</td>
<td>0.01</td>
<td>.23</td>
<td>1.56</td>
<td>.125</td>
</tr>
<tr>
<td>Number of misconceptions</td>
<td>-0.01</td>
<td>0.01</td>
<td>-.10</td>
<td>-.69</td>
<td>.493</td>
</tr>
</tbody>
</table>

$R^2 = .07$, $F(2, 44) = 1.68$, $p = .198$

| **Prediction of Procedural Comprehension** |       |     |         |         |      |
| Constant                      | 0.43  | 0.12| 3.61    | .001    |      |
| Reading skills                | 0.01  | 0.01| .20     | 1.33    | .192 |
| Number of misconceptions      | 0.02  | 0.02| .15     | 1.00    | .322 |

$R^2 = .05$, $F(2, 44) = 1.20$, $p = .312$
Table 5

Regression Analyses for Postdiction Magnitudes

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$b$</th>
<th>SE</th>
<th>$\beta$</th>
<th>$t(44)^a$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.66</td>
<td>0.09</td>
<td></td>
<td>7.73</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Reading skills</td>
<td>0.01</td>
<td>0.01</td>
<td>0.33</td>
<td>2.52</td>
<td>0.015</td>
</tr>
<tr>
<td>Number of misconceptions</td>
<td>-0.03</td>
<td>0.01</td>
<td>-0.35</td>
<td>-2.63</td>
<td>0.012</td>
</tr>
</tbody>
</table>

$R^2 = .27$, $F(2, 44) = 7.94$, $p = .001$

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$b$</th>
<th>SE</th>
<th>$\beta$</th>
<th>$t(44)^a$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.54</td>
<td>0.17</td>
<td></td>
<td>3.16</td>
<td>.003</td>
</tr>
<tr>
<td>Reading skills</td>
<td>0.01</td>
<td>0.01</td>
<td>0.27</td>
<td>1.87</td>
<td>.069</td>
</tr>
<tr>
<td>Number of misconceptions</td>
<td>-0.02</td>
<td>0.02</td>
<td>-0.11</td>
<td>-0.78</td>
<td>.442</td>
</tr>
</tbody>
</table>

$R^2 = .10$, $F(2, 43) = 2.33$, $p = .110$

$^a$As there was a missing value for the postdiction of procedural comprehension, $df = 43$ for the respective analysis.
Table 6  

Regression Analyses for Prediction Biases

<table>
<thead>
<tr>
<th>Predictor</th>
<th>b</th>
<th>SE</th>
<th>β</th>
<th>t(44)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prediction Bias of Conceptual Comprehension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.10</td>
<td>0.10</td>
<td>-0.92</td>
<td>-0.92</td>
<td>.362</td>
</tr>
<tr>
<td>Reading skills</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.03</td>
<td>-0.23</td>
<td>.821</td>
</tr>
<tr>
<td>Number of misconceptions</td>
<td>0.03</td>
<td>0.01</td>
<td>0.26</td>
<td>1.76</td>
<td>.043</td>
</tr>
<tr>
<td>R^2 = .07, F(2, 44) = 1.69, p = .197</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Prediction Bias of Procedural Comprehension** |     |     |     |       |    |
| Constant                           | -0.08 | 0.21 | -0.36 | -0.36 | .718 |
| Reading skills                     | -0.01 | 0.01 | -0.15 | -1.04 | .303 |
| Number of misconceptions           | 0.06  | 0.03 | 0.27 | 1.91  | .032 |
| R^2 = .11, F(2, 44) = 2.75, p = .075 |
Table 7

Regression Analyses for Postdiction Biases

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$b$</th>
<th>SE</th>
<th>$\beta$</th>
<th>$t(44)^a$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Postdiction Bias of Conceptual Comprehension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.08</td>
<td>0.11</td>
<td>-0.74</td>
<td>0.465</td>
<td></td>
</tr>
<tr>
<td>Reading skills</td>
<td>0.01</td>
<td>0.01</td>
<td>0.09</td>
<td>0.59</td>
<td>0.559</td>
</tr>
<tr>
<td>Number of misconceptions</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.12</td>
<td>0.902</td>
</tr>
</tbody>
</table>

$R^2 = .01$, $F(2, 44) = 0.17$, $p = .841$

| **Postdiction Bias of Procedural Comprehension** |       |     |         |           |       |
| Constant                         | 0.03  | 0.14| 0.25    | 0.806     |       |
| Reading skills                   | -0.01 | 0.01| -0.07   | -0.43     | 0.672 |
| Number of misconceptions         | 0.02  | 0.02| 0.16    | 1.04      | 0.305 |

$R^2 = .03$, $F(2, 43) = 0.72$, $p = .493$

$^a$As there was a missing value for the postdiction of procedural comprehension, $df = 43$ for the analysis on postdiction bias of procedural comprehension.
Appendix A

Text¹

Covariance

Covariance is a measure of association that is used to describe linear relationships. A relationship between two variables is linear if it can be represented by a fit line. If the pairs of observations on two variables are drawn as points in a scatter plot with the axes representing the $x$- and $y$-variables, the points would have to lie exactly on a line when there is a linear relationship. However, for nearly all variable relationships that are of interest in research, this requirement is not met. Nevertheless, the line is a useful model to describe the linear relationship between two variables. For example, the trend between body weight $x$ and body height $y$ for a sample of people could be linearly described even if not all points lie exactly on a line. Thereby, the height of the covariance reflects how strongly the points spread around the hypothetical line. If the absolute covariance value is high, this means that the points lie close to the line. The slope of the line, however, depends on the measuring units in which the $x$- and $y$-variables are represented in the scatter plot.

One receives a high positive covariance if frequently an above-average value in variable $x$ corresponds to an above-average value in variable $y$ and, accordingly, if frequently a below-average value in variable $x$ corresponds to a below-average value in variable $y$.

Figure 1. Graphical illustration of positive covariance.

There will be a high negative covariance if frequently an above-average value in variable $x$ corresponds to a below-average value in variable $y$ and, accordingly, if frequently a below-average value in variable $x$ corresponds to an above-average value in variable $y$.

---

¹From *Statistik für Human- und Sozialwissenschaftler* [Statistics for humanities and social science researchers] (pp. 153-156), by J. Bortz and C. Schuster, 2010, Berlin, Germany: Springer. Copyright 2010 by Springer. Translated from German and adapted with permission.
If there is a covariance of zero between two variables, for above-average values in variable $x$ there will be above-average values in variable $y$ as well as below-average values in variable $y$ and vice versa. Therefore, a covariance of zero indicates that there is no linear relationship between two variables. In this case, however, the variables can still be related in a nonlinear manner. Conceivable relationships are, for example, exponential, logarithmic, S-shaped, or U-shaped relationships.

The covariance between two variables can be calculated with the following formula:

$$s_{xy} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

Every studied object produces a pair of observation $(x, y)$, whereby $x$ and $y$ can lie more or less far away from the respective sample mean. If both values lie well above or well below the mean, there is a high positive product of deviation $(x_i - \bar{x}) \cdot (y_i - \bar{y})$. If the values do not deviate much, the product of deviation will be smaller. The sum of the products of deviation of all objects is therefore a measure of the degree of “covariation” of the series of pairs of observation. To account for the number of objects that are included in this sum, the sum is divided by the sample size $N$.

Therefore, covariance indicates if two variables vary together. Concerning the question if two variables are causally related, covariance does not provide statistical evidence. However, covariance can be used to get a first hint towards a potential cause-and-effect relationship.

Moreover, as a nonstandardized statistic, covariance is dependent on the measuring units of the underlying variables. Theoretically, covariance values can reach from negative infinity to positive infinity.
For example, if in two studies the variables body weight and/or body height are measured in different units (e.g., weight: kg, g; height: m, cm), the studies will yield covariances between these two variables that are not directly comparable.
Appendix B

Misconceptions Test

Question 1: *(Misconception: Covariance enables definite predictions)*

In a survey in a service company, the number of years of employment and the job satisfaction (scale from 0 = *very dissatisfied* to 20 = *very satisfied*) of all employees were assessed. The data analysis showed that there was a positive covariance between the two variables. Which of the following conclusions with regard to this result is correct?

- ☐ The job satisfaction of older employees is greater than the job satisfaction of younger employees. *(incorrect)*
- ☐ If it is known for how many years an employee is employed, it is possible to exactly predict his job satisfaction. *(misconception)*
- ☐ To increase job satisfaction, it is beneficial to draw employment contracts for a longer period of time. *(incorrect)*
- ☐ The shorter an employee is employed, the more dissatisfied with his job he tends to be. *(correct)*

Question 2: *(Misconception: Negative covariance indicates the absence of a relationship)*

In classes 8a and 8b of a school, the relationship between reading speed (in seconds) and text comprehension (0 to 10 points in a test) was tested. In class 8a the covariance was 4.50, whereas in class 8b the covariance was -4.50. Which result indicates a relationship between reading speed and text comprehension?

- ☐ None of the two results because one cannot calculate covariance from speed data. *(incorrect)*
- ☐ The result in class 8a because the covariance is positive. *(misconception)*
- ☐ Both results because the absolute covariance value is bigger than zero in both classes. *(correct)*
- ☐ The result in class 8b because the covariance is negative. *(incorrect)*

---

2^Parts added for clarity appear in italics. These parts were not provided to the participants in the study.
**Question 3:** *(Misconception: Covariance is related to the slope of the fit line)*

In multiple studies, a group of doctors calculated the covariance between smoking of cigarettes (average cigarette consumption per day) and severity of a chronic lung disease (scale from 0 = no symptoms to 12 = severe symptoms). In the scatter plots, the results of three of their studies are depicted. According to the scatter plots, which study revealed the highest covariance?

- ☐ Study C because the cloud of points is the densest. *(incorrect)*
- ☐ All studies found the same covariance because there is the same number of points in each scatter plot. *(incorrect)*
- ☐ Study B because the points deviate least from the line. *(correct)*
- ☐ Study A because the slope of the line is the steepest. *(misconception)*
**Question 4:** (Misconception: A straight fit line without deviation invariably indicates perfect covariance)

A cognitive scientist who investigated the dependency of fluid intelligence and creative thinking conducted tests with 10 university students to assess these two variables. The data obtained are drawn in the scatter plot. Which statement concerning the covariance between fluid intelligence and creative thinking can be made based on the scatter plot?

- ☐ There is a negative covariance because the values in creative thinking do not raise with increasing fluid intelligence. *(incorrect)*
- ☐ There is a covariance of zero because all values in creative thinking are identical. *(correct)*
- ☐ There is a positive covariance because the scale of fluid intelligence just starts at a value of 90. *(incorrect)*
- ☐ There is a perfect covariance because all points lie exactly on a line. *(misconception)*

**Question 5:** (Misconception: Covariance implies causality)

In a study, 150 students of an upper secondary school kept a daily record of the number of hours they watched television over a period of one month. Based on their record cards, each students’ average academic achievement (0 = poor to 100 = excellent) was determined. Analysis of the data revealed a negative covariance between television consumption and academic achievement. Which statement concerning this result is true?

- ☐ The reason for the negative covariance probably is that the students had to record their daily television consumption themselves. *(incorrect)*
- ☐ If a student reduces television consumption, this leads to better academic achievement. *(misconception)*
- ☐ One cannot conclude that greater television consumption causes worse academic achievement. *(correct)*
- ☐ A month is not a sufficiently long period to determine students’ average television consumption per day and measure its covariance with other variables. *(incorrect)*
Question 6: (Misconception: Positive covariance is always stronger than negative covariance)

The Federal Institute for Sports Science and Health determined the relation between body height, in meter, and body weight, in kilogram, in four German cities. The following covariances were obtained in the cities:

- Dresden: 1.13
- Cologne: -2.04
- Stuttgart: 0.03
- Hamburg: 1.74

In which city was the strongest covariance between body height and body weight found?

☐ In Hamburg because the covariance with the highest positive value was found there. *(misconception)*

☐ In Stuttgart because the covariance was close to zero there. *(incorrect)*

☐ In Cologne because the covariance with the highest absolute value was found there. *(correct)*

☐ Concerning this matter, no statement can be made because the cities are of different size. *(incorrect)*

Question 7: (Misconception: Covariance can be interpreted from isolated points in scatter plots)

In a study conducted by the German Federal Labor Market Authority, it was examined if there is a relationship between years of vocational training and monthly income (gross). The obtained data were plotted in a graph. Based on the scatter plot, which statement concerning the covariance between years of vocational training and income per month can be made?

☐ There is a negative covariance because people who need longer for their vocational training automatically earn less money. *(incorrect)*

☐ There is a positive covariance because an above-average number of vocational training years frequently corresponds to an above-average income per month. *(correct)*

☐ There is no covariance because for the lowest as well for the highest number of vocational training years the same value for income per month occurs. *(misconception)*

☐ No statement can be made because it is unclear how the scatter plot would look like if there were data for a vocational training time of more than 7 years available. *(incorrect)*
Question 8: *(Misconception: Zero covariance indicates the absence of any association)*

In a study with 40 students of secondary education, students’ reading skill as well as their problem-solving ability were assessed to determine the relationship between these two competencies. The analysis showed that the covariance was zero. Which of the following statements concerning this result is correct?

- Reading skill and problem-solving ability should be trained separately in secondary education. *(incorrect)*
- Probably there is a nonlinear relationship between reading skill and problem-solving ability. *(correct)*
- There has to be an error in the analysis because it is impossible to receive a covariance of zero with a sample of 40 students. *(incorrect)*
- Reading skill and problem-solving ability are related in no manner. *(misconception)*

Question 9: *(Misconception: Only consider covariance if attributable to a causal relationship)*

In the context of a project competition, two judges, judge A and judge B, independently assigned 20 participants a rank from 1 to 20. It was revealed that the covariance between the ranking of judge A and the ranking of judge B was positive. How can this result be interpreted?

- Judge A and judge B did not assign their ranks objectively. *(incorrect)*
- Due to the low number of participants, no conclusion concerning the relationship between the two rankings is warranted. *(incorrect)*
- There is no relationship between the two rankings because the ranking of one judge cannot influence the ranking of the other judge. *(misconception)*
- A lower rank of judge B tends to go along with a lower rank of judge A. *(correct)*
**Question 10:** *(Misconception: Only a straight fit line without deviation indicates covariance)*

In several studies, psychologists analyzed the covariance between the personality traits neuroticism (scale from 0 to 10) and conscientiousness (scale from 0 to 10). The scatter plots show the results of the studies. Which scatter plot indicates that there is covariance between the two variables?

- ☐ All scatter plots because each swarm of points can be easily represented by a line. *(correct)*
- ☐ Scatter plot a and scatter plot b because there are linear deviations. *(incorrect)*
- ☐ Scatter plot c because all points lie exactly on a line. *(misconception)*
- ☐ None of the scatter plots because all lines have fixed start and end points. *(incorrect)*
**Question 11:** *(Misconception: Positive and zero covariance are always stronger than negative covariance)*

In studies by the EU Health Committee, the relationship between income (€ per working hour) and health-related life satisfaction (scale from 0 to 12) in various EU countries was investigated. The following covariances were obtained in four countries:

- Spain: 2.67
- Lithuania: -3.37
- Slovenia: 1.30
- Germany: 0.02

In which line are the covariances correctly ordered from the strongest to the weakest relationship?

☐ Germany: 0.02, Slovenia: 1.30, Spain: 2.67, Lithuania: -3.37 *(incorrect)*

☐ Lithuania: -3.37, Germany: 0.02, Slovenia: 1.30, Spain: 2.67 *(incorrect)*

☐ Lithuania: -3.37, Spain: 2.67, Slovenia: 1.30, Germany: 0.02 *(correct)*

☐ Spain: 2.67, Slovenia: 1.30, Germany: 0.02, Lithuania: -3.37 *(misconception)*

**Question 12:** *(Misconception: Covariance is a standardized statistic)*

In a study, sports scientists from a university determined the covariance between the height and the time for a 100-m race of 20 sprinters. In his calculation, sports scientist A quantified time in seconds. When his colleague, sports scientist B, checked again, he quantified time in milliseconds. Which of the following statements about the covariances calculated by the two sports scientists is correct?

☐ Sports scientist B obtained a higher covariance than sports scientist A because milliseconds yield bigger numbers than seconds. *(correct)*

☐ Both calculations yielded no covariance because one cannot calculate covariance from time data. *(incorrect)*

☐ No statement about the two covariances can be made because it is unknown whether the time and height variables were linear. *(incorrect)*

☐ The two sports scientists received the same covariance because it does not matter whether sports scientist B used milliseconds and sports scientist A used seconds for the calculation. *(misconception)*
Question 13: (Misconception: Covariance changes if the variables are reversed)

In a study by the Leibniz Institute for Science and Mathematics Education, a dependency between the attitude of students towards mathematics and their achievement in mathematics was investigated. Between attitude towards mathematics and grade in mathematics there was a covariance of 6.55. Which result would this study yield if one would reversely calculate the covariance between grade in mathematics and attitude towards mathematics?

☐ The study would yield a covariance of 6.55. (correct)
☐ The study would yield a covariance of -6.55. (misconception)
☐ To make a statement, one would have to know the variance of the sample values. (incorrect)
☐ To make a statement, one would have to know the sample size. (incorrect)

Question 14: (Misconception: More negative covariance indicates a weaker relationship)

A health institute in Munich investigated the relationship between daily fruit consumption and body weight, both measured in kilogram. The results revealed a covariance of -6.23. Likewise, a health center in Berlin analyzed the relationship between fruit consumption per day and body weight. They also used kilograms as measuring units. Here, a covariance of -2.56 was revealed. Which of the two studies revealed the weaker relationship?

☐ The study by the health institute in Munich because the covariance was more negative than in the study by the health center in Berlin. (misconception)
☐ The study by the health center in Berlin because the absolute value of the covariance was smaller than in the study by the health institute in Munich. (correct)
☐ The strength of the relationship does not differ between the two studies because both covariances are smaller than zero. (incorrect)
☐ Concerning this question, no statement can be made because the two studies did not use the same participants. (incorrect)
**Question 15:** *(Misconception: The fit line of perfect covariance must have a constant slope of 45 degrees)*

In two studies, an intelligence researcher investigated the relationship between ability of spatial thinking (scale from 0 to 10) and logical-mathematical intelligence (scale from 0 to 8). The scatter plots show the results of the studies. Which study revealed a perfect covariance between the two variables?

- ☐ Study 1 because the slope of the line has exactly 45°. *(misconception)*
- ☐ None of the two studies because the covariance in both cases is zero. *(incorrect)*
- ☐ Study 2 because the slope of the line has less than 45°. *(incorrect)*
- ☐ Both studies because in both cases the points lie exactly on a line. *(correct)*
## Appendix C

### Misconceptions About Covariance and Respective References

<table>
<thead>
<tr>
<th>Misconception</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1. Covariance enables definite predictions</td>
<td>Batanero et al. (1997); Batanero et al. (1996); Batanero et al. (1998); Estepa and Batanero (1996); Estepa et al. (1999); Liu (2010); Liu and Lin (2008, January); Liu et al. (2010)</td>
</tr>
<tr>
<td>M2. Negative covariance indicates the absence of a relationship</td>
<td>Batanero et al. (1997); Batanero et al. (1996); Batanero et al. (1998); Liu and Lin (2008, January); Mevarech and Kramarsky (1997); Moritz (2004); Morris (1998); Morris (1999); Morris (2001)</td>
</tr>
<tr>
<td>M3. Covariance is related to the slope of the fit line</td>
<td>Liu (2010); Liu and Lin (2008, January); Liu et al. (2010); Liu et al. (2009)</td>
</tr>
<tr>
<td>M4. A straight fit line without deviation invariably indicates perfect covariance</td>
<td>Liu (2010); Liu and Lin (2008, January); Liu et al. (2009)</td>
</tr>
<tr>
<td>M5. Covariance implies causality</td>
<td>Estepa and Sánchez (2001); Garfield (2003); Liu and Lin (2008, January); Liu et al. (2009); Morris (1998); Morris (1999); Morris (2001); Sundre (2003, April)</td>
</tr>
<tr>
<td>M6. Positive covariance is always stronger than negative covariance</td>
<td>Estepa and Sánchez (2001); Liu (2010); Liu and Lin (2008, January); Liu et al. (2010); Liu et al. (2009); Morris (1998); Morris (1999); Morris (2001)</td>
</tr>
<tr>
<td>M7. Covariance can be interpreted from isolated points in scatter plots</td>
<td>Batanero et al. (1997); Batanero et al. (1996); Batanero et al. (1998); Estepa and Batanero (1996); Estepa et al. (1999); Mevarech and Kramarsky (1997); Moritz (2004)</td>
</tr>
</tbody>
</table>
| M8. Zero covariance | Estepa and Sánchez (2001); Liu (2010); Liu and Lin (2008, January); Liu et al. (2010); Liu et al. (2009) 
indicates the absence of any association |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M9. Only consider covariance if attributable to a causal relationship</td>
<td>Batanero et al. (1997); Batanero et al. (1998); Estepa and Batanero (1996)</td>
</tr>
<tr>
<td>M10. Only a straight fit line without deviation indicates covariance</td>
<td>Liu and Lin (2008, January); Liu et al. (2009)</td>
</tr>
<tr>
<td>M11. Positive and zero covariance are always stronger than negative covariance</td>
<td>Liu et al. (2009); Morris (1999)</td>
</tr>
<tr>
<td>M12. Covariance is a standardized statistic</td>
<td>Batanero et al. (1997); Batanero et al. (1998); Estepa and Sánchez (2001); Liu (2010); Liu and Lin (2008, January); Liu et al. (2010)</td>
</tr>
<tr>
<td>M13. Covariance changes if the variables are reversed</td>
<td>Estepa and Sánchez (2001); Liu (2010); Liu and Lin (2008, January); Liu et al. (2010)</td>
</tr>
<tr>
<td>M14. More negative covariance indicates a weaker relationship</td>
<td>Batanero et al. (1997); Batanero et al. (1998); Liu et al. (2009)</td>
</tr>
<tr>
<td>M15. The fit line of perfect covariance must have a constant slope of 45 degrees</td>
<td>Liu and Lin (2008, January); Liu et al. (2009)</td>
</tr>
</tbody>
</table>
Appendix D

Conceptual Comprehension Questions³

Question 1:
In two samples, sample A and sample B, the covariance between variable x and variable y has been calculated. In both samples, the covariance is positive. However, the covariance in sample A is higher than in sample B. What could be the reason for the higher covariance in sample A compared with sample B?

☐ In sample A more above-average values in variable x go along with below-average values in variable y than in sample B. (incorrect)

☐ In sample A more above-average values in variable x go along with above-average values in variable y than in sample B. (correct)

☐ In sample A less below-average values in variable x go along with below-average values in variable y than in sample B. (incorrect)

☐ In sample A more below-average values in variable x go along with above-average values in variable y than in sample B. (incorrect)

Question 2: (Misconception: Negative covariance indicates the absence of a relationship)
In classes 8a and 8b of a school, the relationship between reading speed (in seconds) and text comprehension (0 to 10 points in a test) was tested. In class 8a the covariance was 4.50, whereas in class 8b the covariance was -4.50. Which result indicates a relationship between reading speed and text comprehension?

☐ None of the two results because one cannot calculate covariance from speed data. (incorrect)

☐ The result in class 8b because the covariance is negative. (incorrect)

☐ The result in class 8a because the covariance is positiv. (misconception)

☐ Both results because the absolute covariance value is bigger than zero in both classes. (correct)

³Parts added for clarity appear in italics. These parts were not provided to the participants in the study.
Question 3:

What does the following scatter plot show about the covariance between the number of hours a student of a class learned for a test and the number of incorrect answers he or she gave in the test?

☐ The covariance between learning time and number of incorrect answers is positive. *(incorrect)*

☐ The covariance between learning time and number of incorrect answers is negative. *(incorrect)*

☐ The covariance between learning time and number of incorrect answers is zero or close to zero. *(correct)*

☐ The covariance between learning time and number of incorrect answers is not possible to determine. *(incorrect)*
Question 4: *(Misconception: Covariance is a standardized statistic)*

In a study, sports scientists from a university determined the covariance between the height and the time for a 100-m race of 20 sprinters. In his calculation, sports scientist A quantified time in seconds. When his colleague, sports scientist B, checked again, he quantified time in milliseconds. Which of the following statements about the covariances calculated by the two sports scientists is correct?

☐ The two sports scientists received the same covariance because it does not matter whether sports scientist B used milliseconds and sports scientist A used seconds for the calculation.  *(misconception)*

☐ No statement about the two covariances can be made because it is unknown whether the time and height variables were linear.  *(incorrect)*

☐ Sports scientist B obtained a higher covariance than sports scientist A because milliseconds yield bigger numbers than seconds.  *(correct)*

☐ Both calculations yielded no covariance because one cannot calculate covariance from time data.  *(incorrect)*

Question 5:

In a study, the relationship between interest in natural sciences (scale from 0 = *very low interest* to 10 = *very high interest*) and achievement in physics (grade from 1 = *very good* to 6 = *insufficient*) was investigated. The results revealed a positive covariance. How can this covariance be interpreted?

☐ Interest in natural sciences and achievement in physics are not related.  *(incorrect)*

☐ When interest in natural sciences is lower, achievement in physics is typically poorer.  *(incorrect)*

☐ When interest in natural sciences is higher, achievement in physics is typically greater.  *(incorrect)*

☐ When interest in natural sciences is higher, achievement in physics is typically poorer.  *(correct)*
**Question 6:** *(Misconception: Covariance is related to the slope of the fit line)*

In multiple studies, a group of doctors calculated the covariance between smoking of cigarettes (average cigarette consumption per day) and severity of a chronic lung disease (scale from 0 = *no symptoms* to 12 = *severe symptoms*). In the scatter plots, the results of three of their studies are depicted. According to the scatter plots, which study revealed the highest covariance?

- ☐ All studies found the same covariance because there is the same number of points in each scatter plot. *(incorrect)*
- ☐ Study A because the slope of the line is the steepest. *(misconception)*
- ☐ Study B because the points deviate least from the line. *(correct)*
- ☐ Study C because the cloud of points is the densest. *(incorrect)*
Question 7:
When does the covariance between power of concentration and learning outcome take on a value of zero?

☐ If for one of the two variables all sample values are identical. (correct)
☐ If for one of the two variables the mean is zero. (incorrect)
☐ If for one of the two variables the variance is 1. (incorrect)
☐ If for one of the two variables the variation is lower than for the other variable. (incorrect)

Question 8: (Misconception: Covariance implies causality)
In a study, 150 students of an upper secondary school kept a daily record of the number of hours they watched television over a period of one month. Based on their record cards, each students’ average academic achievement (0 = poor to 100 = excellent) was determined. Analysis of the data revealed a negative covariance between television consumption and academic achievement. Which statement concerning this result is true?

☐ One cannot conclude that greater television consumption causes worse academic achievement. (correct)
☐ The reason for the negative covariance probably is that the students had to record their daily television consumption themselves. (incorrect)
☐ A month is not a sufficiently long period to determine students’ average television consumption per day and measure its covariance with other variables. (incorrect)
☐ If a student reduces television consumption, this leads to better academic achievement. (misconception)
Appendix E

Procedural Comprehension Questions

Question 1:

In a study, the following data for variable $x$ and variable $y$ were obtained for 4 participants:

<table>
<thead>
<tr>
<th>Participant</th>
<th>Variable $x$</th>
<th>Variable $y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

Use these data to calculate the covariance between the two variables.
Question 2:

For variable x and variable y the following data were obtained for 5 participants:

<table>
<thead>
<tr>
<th>Participant</th>
<th>Variable x</th>
<th>Variable y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Create a scatter plot in which you draw in the data and indicate whether the covariance is positive, negative, or zero.
Question 3:

Two variables were assessed on 5 participants. The data obtained are drawn in the scatter plot. What is the covariance between the two variables? Provide a calculation or an explanation.
Question 4:

In a study, the covariance between two variables was determined, and the analysis revealed a value of 200. However, erroneously, when calculating the covariance, only the sum of the products of deviation was calculated, and the sample size with $N = 20$ was forgotten. Considering this, what is the actual covariance between the two variables?
Appendix F

Results on Metacomprehension Accuracy in Terms of Absolute Accuracy

Misconceptions and Prediction Accuracy

As can be seen in Table F1, the linear regression analyses revealed that misconceptions neither significantly influenced the prediction absolute accuracy of conceptual comprehension nor the prediction absolute accuracy of procedural comprehension. Hence, in contrast to bias, the absolute accuracy with which participants predicted their conceptual and procedural text comprehension was not affected by their misconceptions. This is because the range of bias was made up of negative and positive values. More precisely, participants with a lower number of misconceptions were rather underconfident, whereas participants with a higher number of misconceptions tended to be overconfident. Thus, while there was an effect of misconceptions on the direction of the deviations between predicted and actual comprehension (bias), there was no effect of misconceptions when only looking at the magnitude of the deviations (absolute accuracy).

Table F1

Regression Analyses for Prediction Absolute Accuracies

<table>
<thead>
<tr>
<th>Predictor</th>
<th>b</th>
<th>SE</th>
<th>β</th>
<th>t(44)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction Absolute Accuracy of Conceptual Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.15</td>
<td>0.07</td>
<td>2.04</td>
<td>.048</td>
<td></td>
</tr>
<tr>
<td>Reading skills</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.02</td>
<td>-0.14</td>
<td>.887</td>
</tr>
<tr>
<td>Number of misconceptions</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.09</td>
<td>.933</td>
</tr>
</tbody>
</table>

\[ R^2 = .01, F(2, 44) = 0.01, p = .988 \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>b</th>
<th>SE</th>
<th>t(44)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction Absolute Accuracy of Procedural Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.42</td>
<td>0.12</td>
<td>3.38</td>
<td>.002</td>
</tr>
</tbody>
</table>
THE DOUBLE CURSE OF MISCONCEPTIONS

<table>
<thead>
<tr>
<th>Predictor</th>
<th>b</th>
<th>SE</th>
<th>β</th>
<th>t(44)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postdiction Absolute Accuracy of Conceptual Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.18</td>
<td>0.06</td>
<td>2.84</td>
<td>.007</td>
<td></td>
</tr>
<tr>
<td>Reading skills</td>
<td>-0.01</td>
<td>0.01</td>
<td>-.16</td>
<td>-1.04</td>
<td>.303</td>
</tr>
<tr>
<td>Number of misconceptions</td>
<td>0.01</td>
<td>0.01</td>
<td>.07</td>
<td>0.44</td>
<td>.659</td>
</tr>
</tbody>
</table>

\[ R^2 = .03, F(2, 44) = 0.74, p = .485 \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>b</th>
<th>SE</th>
<th>β</th>
<th>t(44)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postdiction Absolute Accuracy of Procedural Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.10</td>
<td>0.10</td>
<td>1.01</td>
<td>.317</td>
<td></td>
</tr>
<tr>
<td>Reading skills</td>
<td>0.01</td>
<td>0.01</td>
<td>.06</td>
<td>0.40</td>
<td>.688</td>
</tr>
<tr>
<td>Number of misconceptions</td>
<td>0.01</td>
<td>0.01</td>
<td>.14</td>
<td>0.94</td>
<td>.353</td>
</tr>
</tbody>
</table>

\[ R^2 = .02, F(2, 43) = 0.47, p = .626 \]

\(^{a}\)As there was a missing value for the postdiction of procedural comprehension, \( df = 43 \) for the analysis on postdiction absolute accuracy of procedural comprehension.

Misconceptions and Postdiction Accuracy

As represented in Table F2, the linear regression analyses showed that misconceptions neither significantly influenced the postdiction absolute accuracy of conceptual comprehension nor the postdiction absolute accuracy of procedural comprehension.

Table F2

Regression Analyses for Postdiction Absolute Accuracies

\[ R^2 = .05, F(2, 44) = 1.03, p = .366 \]