High-Performing Readers Underestimate Their Text Comprehension: Artifact or Psychological Reality?

Stefanie Golke (stefanie.golke@ezw.uni-freiburg.de)
Department of Educational Science, University of Freiburg
Rempartstr. 11, 79098 Freiburg, Germany

Jörg Wittwer (joerg.wittwer@ezw.uni-freiburg.de)
Department of Educational Science, University of Freiburg
Rempartstr. 11, 79098 Freiburg, Germany

Abstract

We focused on the controversy whether high-performing readers consistently underestimate their comprehension or are prone to detrimental overestimations as much as less skilled readers are. Therefore, we conducted an experiment (N = 105 university students) to investigate judgment bias as a function of reading skill and text difficulty in terms of text cohesion. Results showed that the easy text produced underestimation of comprehension, whereas the hard text led to overestimation. Furthermore, readers with higher reading skills were less prone to overestimate their comprehension of a hard text than less skilled readers. However, we also found that more skilled readers showed lower sensitivity in discriminating between correct and incorrect answers than less skilled readers. Overall, our results do not support the idea that high-performing readers consistently underestimate their text comprehension. Findings are discussed with respect to readers' awareness of different text-based judgment cues and their (beliefs about their) reading skill.

Keywords: judgment bias; metacognitive sensitivity; text difficulty; reading skill; high-performing readers

Introduction

Successful learning from text requires readers to accurately judge their text comprehension because false judgments (e.g., overestimation) can hamper the learning process. It is well acknowledged that readers in general are prone to overestimations, whereas particularly high-performing readers (i.e., readers who achieve high scores on a text comprehension test) might more likely underestimate their comprehension (de Bruin et al., 2016; Dunlosky & Rawson, 2012). However, the methodology used to unveil underestimation by high-performing readers is not fully undisputed. Therefore, it remains unclear whether high-performing readers' underestimation is psychological reality or rather an artifact. We present an experiment that was conducted to advance our understanding about how high-performing readers judge their text comprehension.

Comprehension Judgments and the Learning Process

Learning from text involves constructing a mental representation of the information provided in a text and retrieving the learned information at a later time. The learning process heavily depends on a reader's metacognitive ability to monitor comprehension (i.e., metacomprehension), which is mirrored in the correspondence between a reader's comprehension judgment and actual performance on a comprehension test (Wiley, Griffin, & Thiede, 2005).

Comprehension judgments can occur at different times in the learning process. Accordingly, research uses different types of comprehension judgments (Griffin, Jee, & Wiley, 2009). The first type is the prospective judgment of comprehension that readers make after reading a text to predict how well they will perform on yet unknown test questions about the text. Furthermore, when readers complete test questions, they can use information about their (perceived) performance in answering the test questions to evaluate their comprehension. Thus, the second type of comprehension judgments is readers' confidence in their retrieved answers on single test questions (i.e., response confidence). The third type is the retrospective judgment of comprehension that refers to a whole set of test questions (i.e., how many of the test questions were answered correctly). The three types of judgment are assumed to reflect (slightly) different aspects of metacomprehension but complement each other (Schraw et al., 2014).

When readers make comprehension judgments, they normally use available cues (Koriat, 1997). These cues can arise from the learning material (e.g., text difficulty), a reader's (self-perceived) skills and resources (e.g., prior knowledge, reading ability) and a reader's experiences when reading the text or answering test questions. All types of cues can be useful for precise judgments when they are valid indicators of the required level of comprehension.

To support learning, judgments of comprehension need to be precise because they influence readers' subsequent learning activities. Imprecise judgments, especially overestimations, have a detrimental effect on learning (Dunlosky & Rawson, 2012). For example, overestimation means that readers do not realize that their comprehension of text is worse than they think. Therefore, they might abstain from engaging in remedial activities. In contrast, underestimation might be less problematic for learning but it can hamper learners in allocating their learning time appropriately.

Controversy Over High-Performing Readers' Underestimation

Numerous studies have shown that readers typically provide imprecise judgments. Most readers overestimate their comprehension of text and are overconfident in the correctness of their retrieved information when answering test questions about the text (Dunlosky & Rawson, 2012; Maki et al., 2005).

However, concerning high-performing readers, it is sometimes reported that they tend to underestimate their comprehension (de Bruin et al., 2016; Zabrucky, 2010). This underestimation is often interpreted as a result of specific metacognitive or cognitive processes. For example, high performers are assumed to not give very high judgments of their comprehension to avoid being perceived as arrogant or to have negatively skewed misperceptions of their abilities, both resulting in underestimation (Zabrucky, 2010).

A completely different explanation of this phenomenon refers to a statistical bias of the measure used to unveil overestimation and underestimation (i.e., judgment bias) that becomes relevant when readers' level of performance is determined by their performance in the experimental comprehension test. More specifically, judgment bias uses the signed difference between a reader's prospective or retrospective judgment of comprehension and his/her actual performance on a comprehension test. Therefore, the reader's judgment bias is constrained by his/her performance (Griffin et al., 2009; see also Kruger & Dunning, 1999). That is, readers who achieve the maximum or a very high performance score on a comprehension test (i.e., high-performing readers) are much more likely to show underestimation than readers with lower performance scores. Conversely, readers who have a very low performance score are much more likely to overestimate their comprehension. Furthermore, if the performance-level of readers is determined by their performance on the comprehension test – that is also part of the measure of judgment bias – both measures are statistically dependent on each other and normally show high negative correlations (i.e., higher performance on the comprehension test is associated with lower/more negative scores of judgment bias). Thus, the finding that high-performing readers underestimate their comprehension could also be a statistical artifact and, hence, might not reflect their actual ability to judge comprehension. To disentangle the effect of the level of comprehension on judgment bias, it seems useful to investigate judgment bias as a function of both readers' general reading skill and test/text difficulty.

The Effect of Text Difficulty and Reading Skill

Maki et al. (2005) investigated judgment bias (i.e., overestimation or underestimation) as a function of text difficulty – determined by the readability of the texts – and students' general reading skill. Their findings did not support the view that high performers generally underestimate their comprehension. Instead, for difficult texts (i.e., lower readability), it was found that high-ability readers were precise when making prospective judgments. Only when making postdictions, they underestimated their comprehension but so did medium-ability readers as well. Conversely, for easier texts (i.e., higher readability but still in the range between difficult and standard texts), all readers provided overoptimistic predictions of comprehension but precise postdictions.

This latter finding on easier texts is intriguing with regard to Schraw and Roedel's (1994) study that determined difficulty by the mean item difficulty of the test questions. They found that readers were overconfident on their answers in response to difficult and moderately difficult items but precise on items with low difficulty. Because high-ability readers in Maki and colleagues' study (2005) solved about 70% of the test items on the easier text, these test items were of low difficulty for them. Hence, their postdiction judgments were precise. But why did the (high-ability) readers overestimate their comprehension when making prospective judgments on the easier text? It appears as if the higher readability of the text might have induced readers – at any level of reading ability – to be overoptimistic. This interpretation is supported by findings from Weaver and Bryant (1995) who revealed that predictions of comprehension are not highly correlated to actual performance for texts with high or low readability.

Thus, previous studies showed that high-performing readers do not consistently underestimate their comprehension. Therefore, these studies provide useful hints about the controversy on high-performing readers. However, at the same time, the studies only provide information about the effects of item difficulty (Schraw & Roedel, 1994) or text difficulty in terms of readability (Maki et al., 2005; Weaver & Bryant, 1995). Readability that depends on, for example, word length, number of words per sentence, or passive/active structure is a salient text-based cue and a more distal indicator of the difficulty of the text content than, for example, text cohesion. With regard to theories on text comprehension (see e.g., Wiley et al., 2005), varying text difficulty in terms of readability might not discriminate well enough between readers with different levels of reading proficiency. Hence, it would be interesting to focus on cohesion as a different indicator of text difficulty and investigate judgment bias as a function of this text feature and reading skill.

The Present Study

We examined the precision of comprehension judgments as a function of text difficulty and reading skill. In contrast to previous studies, we determined text difficulty in terms of text cohesion. To assess judgment bias, we used the signed difference between a reader's prospective or retrospective judgment of comprehension and his/her actual performance on a comprehension test. Moreover, we assessed readers'

metacognitive sensitivity and response bias as additional indicators of metacomprehension.

As main effects of text difficulty on performance, we expected that the easy text resulted in higher performance on the test questions than the hard text. Regarding the effect of text difficulty on judgment bias, we based our hypotheses on findings about item difficulty instead of texts' readability. Therefore, given the statistical dependence of performance level and judgment bias, we hypothesized that the easy text would lead to significant underestimation whereas the hard text should result in significant overestimation. Hence, using a within contrast, the easy text should result in a lower bias score of prospective and retrospective judgments than the hard text. Furthermore, we investigated in an exploratory way how reading skill was linked to readers' judgment bias for the easy text compared with the hard text. To do so, we computed multiple linear regressions that included reading skill and prior knowledge as relevant predictors of prospective judgment bias for the easy text and the hard text. In case of the retrospective judgment bias, we also used readers' metacognitive sensitivity and response bias that are based on readers' response confidence for the test questions as additional predictors. With regard to the relationship of reading skill with metacognitive sensitivity and response bias, we inspected their correlations with each other.

Method

Design

The experiment followed a two-factorial design with reading skill as a metric between-subjects factor and text difficulty as the within-subjects factor with two levels: one text with lower text difficulty (easy text) and one text with higher text difficulty (hard text) in terms of cohesive relations within the text (see also Materials). The order of the texts was counterbalanced across all participants.

As dependent variables, we assessed: 1) text comprehension (i.e., number of correctly answered questions about the text), 2) the bias of prospective and retrospective judgments, 3) metacognitive sensitivity, and 4) response bias. Furthermore, we assessed participants' prior knowledge about the topics of the text materials.

Participants

Participants were 105 university students from educational science. They had a mean age of 22.78 (*SD* = 4.95) years and 82% of them were female.

Materials

Table 1 displays the main characteristics of both texts. Given the scope of this study, we selected texts that represented different levels of text difficulty in terms of cohesion. Cohesion refers to the extent to which relations between ideas in a text are made explicit by using, for example, textual features such as causal, temporal, or additive connectives. We determined cohesion by the proportion of sentences that contained a cohesive device on how the sentence is connected to previous ones. As displayed in Table 1, the cohesion score for the hard text was considerably lower than the score for the easy text. Thus, the hard text required readers to engage more deeply in comprehending the text compared with the easy text. Apart from cohesion, the texts were equivalent with respect to other characteristics including surface cues, such as readability or text length, as well as the domain of the texts (i.e., biology, see Table 1).

We used six open-ended comprehension questions for each text. The questions tapped information explicitly stated in the text.

Table 1: Characteristics of the texts.

Characteristic	Easy text	Hard text
Topic	Reproduction	Immunology
No. of words	380	397
No. of sentences	25	30
Flesch-Index[a]	46	41
Cohesion	0.67	0.38

Note. [a]Texts with a Flesch-Index (i.e., flesch reading ease score) between 30 and 50 reflect difficult texts in terms of readability that are typically used in higher education.

Instruments and Measures

Prospective and Retrospective Judgments Participants indicated how many of the six text comprehension questions they think they would answer correctly (= prospective judgment) or had answered correctly (= retrospective judgment; value between 0 and 6).

Judgment Bias We used the signed difference between a reader's prospective or retrospective judgment of comprehension and the actual performance on the text comprehension test. Hence, the bias score could range between -6 (i.e., maximum underestimation) and +6 (i.e., maximum overestimation).

Response Confidence For each question, participants indicated how confident they were that their answer was correct (Likert scale from 1 = *very uncertain* to 7 = *very certain*).

Metacognitive Sensitivity (*d´*) Sensitivity reflects the ability of readers to distinguish between correct and incorrect responses on test questions. It uses readers' performance on single test questions and their response confidence on these test questions. We determined metacognitive sensitivity via *d´* that is based on signal detection theory (see Fleming & Lau, 2014; Schraw et al., 2014) using the hit rate (i.e., number of questions that a reader answered correctly and rated as correct, divided by the total number of correctly answered questions) and the false alarm rate (i.e., number of questions that a reader did not answer correctly but rated as correct, divided by the total number of incorrect answers). The measure of *d´* is the difference between the standardized hit

rate and the standardized false alarm rate. A value of zero means that the reader could not discriminate between correct and incorrect responses, a positive value (i.e., higher hit rate than false alarm rate) reflects good sensitivity, and a negative value (i.e., higher false alarm rate than hit rate) suggests that the reader considered rather a false answer as correct than a correct answer.

Response Bias (*c*) The response bias *c* is based on the sensitivity measure *d*' [*c* = -0.5 * (standardized hit rate + standardized false alarm rate)]. The response bias represents the tendency of a reader to accept false alarms (*c* < 0) or to be cautious when giving confidence judgments on single test questions in order to avoid false alarms (*c* > 0).

Reading Skill We used a subtest of a computer-based German reading comprehension test for adults (ELVES; Richter & van Holt, 2005). The subtest assessed higher-order processes of text comprehension.

Prior Knowledge There was a total of 12 open-ended questions that assessed readers' prior knowledge on immunology and reproduction. These questions were not identical to the text comprehension questions.

Procedure

At the beginning, participants answered the prior knowledge test and proceeded with the reading comprehension test ELVES. After that, participants read the first experimental text and then judged their comprehension by predicting how many of the six text comprehension questions they think they would answer correctly. After the judgment, they answered the comprehension questions and rated their response confidence for each question. After answering all comprehension questions, participants made a retrospective comprehension judgment by indicating how many of the six questions they thought they had answered correctly. Subsequently, participants proceeded with the second experimental text in the same manner as they did for the first one.

Results

To test the hypotheses regarding the main effect of text difficulty on performance and judgment bias, we performed (paired) *t*-tests (for descriptive statistics, see Table 2). In line with our hypotheses, we found that the easy text resulted in higher performance on the text comprehension questions, $t(104) = 13.73$, $p < .001$, Cohens $d = 1.49$ (large effect), than the hard text. Moreover, the mean scores of prospective and retrospective judgment bias for both texts (see Table 2) were significantly different from zero (i.e., the value of perfect judgment), all *p*'s < .004. Thus, the easy text resulted in significant underestimation for both prospective and retrospective judgments. In contrast, the hard text resulted in significant overestimation for both types of judgment. A paired *t*-test confirmed that the easy text resulted in lower bias scores of prospective judgments, $t(104) = -12.96$, $p < .001$, Cohens $d = -1.42$ (large effect), and lower bias scores of retrospective judgments, $t(104) = -6.13$, $p < .001$, Cohens $d = -0.68$ (medium effect), than the hard text.

Furthermore, we performed multiple linear regressions to examine our research question regarding the relationship of reading skill with judgment bias for the easy and the hard text. For each type of judgment bias (i.e., prospective vs. retrospective bias), we computed separate multiple regressions for the easy and the hard text. Predictors were entered in one step.

Table 2: Means (and standard deviations) for dependent variables as a function of text difficulty.

Dependent variable	Easy text	Hard text
Text comprehension	4.90 (1.31)	2.99 (1.25)
Prospective judgment bias	-0.79 (1.42)	1.21 (1.40)
Retrospective judgment bias	-0.35 (1.18)	0.49 (1.29)

Regarding prospective judgment bias, we included prior knowledge on the topic of the text and reading skill as predictors. The results (see Table 3) showed that neither prior knowledge nor reading skill were statistically relevant predictors of prospective judgment bias for the easy text. However, for the hard text, reading skill was a statistically significant negative predictor of prospective judgment bias. That is, participants with higher reading skills were less likely to overestimate their comprehension of the hard text. However, as descriptive statistics revealed (see Table 2), we cannot conclude that these participants generally showed underestimation because only 12% of the total sample underestimated their comprehension of the hard text when making prospective judgments.

Table 3: Predictors of prospective judgment bias for easy and hard text.

Predictor	*b*	*SE b*	*t*(101)	*p*
Easy text				
Constant	-0.39	0.53	-0.74	.462
Reading skill	-0.02	0.03	-0.65	.516
Prior knowledge	0.00	0.01	-0.32	.749
Hard text				
Constant	2.43	0.49	4.96	< .001
Reading skill	-0.08	0.03	-2.83	.006
Prior knowledge	0.00	0.01	-0.15	.881

Note. For easy text: $R^2 = .01$, $F(2, 102) = 0.32$, $p = .730$. For hard text: $R^2 = .08$, $F(2, 102) = 4.16$, $p = .018$.

Moreover, regarding retrospective judgment bias, we included prior knowledge, reading skill as well as

metacognitive sensitivity and response bias as predictors. The multiple regression analyses revealed (see Table 4) that metacognitive sensitivity and response bias significantly predicted the retrospective judgment bias for the easy text. That is, the better a reader discriminated between correct and incorrect responses and the more the readers avoided false alarms in the confidence rating, the less likely this reader was to overestimate comprehension when making retrospective judgments on questions about an easy text. This result was also found for the hard text. Additionally, reading skill also predicted retrospective judgment bias for the hard text.

Table 4: Predictors of retrospective judgment bias for easy and hard text.

Predictor	*b*	*SE b*	*t*(99)	*p*
Easy text				
Constant	-0.02	0.41	-0.05	.958
Reading skill	-0.03	0.02	-1.30	.222
Prior knowledge	0.00	0.01	0.31	.761
Sensitivity	-0.24	0.08	-2.96	.004
Response bias	-0.59	0.14	-4.13	< .001
Hard text				
Constant	1.51	0.46	3.30	.001
Reading skill	-0.06	0.02	-2.34	.022
Prior knowledge	0.00	0.01	-0.48	.632
Sensitivity	-0.16	0.09	-1.78	.078
Response bias	-0.78	0.16	-5.04	< .001

Note. For easy text: $R^2 = .22$, $F(4, 104) = 6.83$, $p < .001$. For hard text: $R^2 = .23$, $F(4, 99) = 7.30$, $p < .001$.

Furthermore, we explored the relationship of reading skill with metacognitive sensitivity and response bias, respectively. As displayed in Table 5, we found that participants with higher reading skills were less cautious (measure of response bias, *c*) when giving confidence ratings on the comprehension questions about the easy text. In addition, they were less able to discriminate between correct and incorrect answers (measure of metacognitive sensitivity, *d´*) in response to questions about the hard text. Given the magnitude of the correlation coefficients, these relations are small effects. However, it appears that more-skilled readers were metacognitively less aware and, therefore, more overconfident when answering the test questions.

Table 5: Pearson's *r* correlations between reading skill, sensitivity (*d´*), and response bias (*c*) for easy and hard text.

	Easy text		Hard text	
Measure	*d´*	*c*	*d´*	*c*
Reading skill	.10	-.22*	-.27**	-.16

Note. $^{*}p < .05$. $^{**}p < .01$.

To sum up, we found that reading skill was a relevant predictor of prospective and retrospective judgment bias in case of the hard text, but not in the case of the easy text. Hence, participants with higher reading skills were less likely to overestimate their comprehension of the hard text. Moreover, we found that response bias and sensitivity influenced retrospective judgment bias for the easy and the hard text. Thus, the better participants discriminated between correct and incorrect answers or the more cautious they were when rating their answers as correct, the less likely they made overoptimistic retrospective judgments. In addition, we found that sensitivity and response bias were more negative for readers with higher reading skills, although these effects were rather small.

Discussion

This study aimed to shed further light on the question whether high-performing readers adhere to judgment processes that lead them to consistently underestimate their comprehension across materials with different levels of difficulty. The results of our study do not support this assumption. Instead, our results suggest that readers with higher reading skills are better calibrated because they are less prone to overestimate their comprehension of a hard text compared with readers with lower reading skills. Kwon and Linderholm (2014) also found this relationship for texts with standard readability.

The finding that participants with higher reading skills were better calibrated supports the notion that higher reading skills include better monitoring during reading. Readers who actively monitor their text comprehension obtain a more comprehensive mental model of the text and are therefore more precise at judging their comprehension (Wiley et al., 2005). Furthermore, although a relationship between reading skill and judgment bias is evident, the magnitude of the relationship we found in our study is rather small. This indicates that other characteristics of the reader are also or even more relevant for judgment bias, for example, the self-perceived reading skill (Kwon & Linderholm, 2014).

In contrast to the hard text, there was no relationship between reading skill and judgment bias on the easy text. This finding can be explained by the low difficulty of the test questions. Therefore, general reading skill was not predictive of test performance on the easy text and, thus, reading skill was not related to judgment bias on the easy text.

Another important finding in our study were the negative relations of reading skill with metacognitive sensitivity and response bias. This finding suggests that readers with higher reading skills may be metacognitively unaware when responding to the type of test questions we used in the present study. Therefore, despite their good calibration with respect to the hard text, participants with higher reading skills showed a flawed discrimination performance. To explain this lower discrimination, it can be speculated that their beliefs about their reading skill tempted high-ability readers to

proceed less mindfully with the test questions and, thus, to be overconfident on their answers.

This interpretation does not necessarily contradict the findings on the positive influence of reading skill and the negative impact of sensitivity and response bias on retrospective judgment bias because the strength of these relations was rather small. Moreover, it can be assumed that other factors influence judgment bias as well. Therefore, the seemingly contradicting relations between reading skill, discrimination performance, and retrospective judgment bias might simply indicate complex interactions between readers' characteristics and judgment processes that still need to be further uncovered (Schraw et al., 2014).

The findings of this study also contribute to the understanding of the effects of text-based cues on judgment bias. In our study, the easy text (i.e., higher cohesion) resulted in underestimation. Given that performance on test questions about the easy text was rather high, this underestimation was very likely to occur due to probabilistic assumptions (Schraw & Roedel, 1994). Likewise, the observed overestimation on the test questions about the hard texts was also expected. In contrast, the easy text (i.e., higher readability) in Maki and colleagues' (2005) study resulted in overestimation of prospective judgments for all readers. Only when readability of texts was low, readers, except for weak readers, adjusted their comprehension judgments. Thus, we can conclude that texts that are easy to read – and, therefore, often preferred in instructional contexts because they increase performance – are more likely to seduce readers to be overoptimistic. Conversely, high text cohesion does not seem to have such an effect on metacognitive judgment. Therefore, readers, including high-ability readers, are apparently unaware of the low validity of good text readability as a cue to judge their comprehension. With respect to readers' sensitivity for text cohesion, we aim to analyze our data in more depth addressing possible anchor effects based on the within-subjects design and also examine the role of reading skill in this regard.

Acknowledgments

This study is based on data collected by Nadine Bianchi, Nina Krämer, and Andreas Leiprecht as part of their unpublished Bachelor theses.

References

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

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction*, *22*, 271–280. doi:10.1016/j.learninstruc.2011.08.003

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 1–9. doi:10.3389/fnhum.2014.00443

Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition*, *37*, 1001–1013. doi:10.3758/MC.37.7.1001

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

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

Kwon, H., & Linderholm, T. (2014). Effects of self-perception of reading skill on absolute accuracy of metacomprehension judgements. *Current Psychology*, *33*, 73–88. doi:10.1007/s12144-013-9198-x

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology, 97*, 723–731. doi:10.1037/0022-0663.97.4.723

Richter, T., & van Holt, N. (2005). ELVES: Ein computergestütztes Diagnostikum zur Erfassung der Effizienz von Teilprozessen des Leseverstehens [A computer-based instrument to assess efficiency of reading processes]. *Diagnostica*, *51*, 169–182.

Schraw, G., Kuch, F., Gutierrez, A. P., & Richmond, A. S. (2014). Exploring a three-level model of calibration accuracy. *Journal of Educational Psychology*, *106*, 1192–1202. doi:10.1037/a0036653

Schraw, G., & Roedel, T. D. (1994). Test difficulty and judgment bias. *Memory & Cognition*, *22*, 63–69. doi:10.3758/BF03202762

Weaver, C. A., & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition*, *23*, 12–22.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology*, *132*, 408–428. doi:10.3200/GENP.132.4.408-428

Zabrucky, K. M. (2010). Knowing what we know and do not know: Educational and real world implications. *Procedia - Social and Behavioral Sciences*, *2*, 1266–1269. doi:10.1016/j.sbspro.2010.03.185